

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 June 2001 (21.06.2001)

PCT

(10) International Publication Number
WO 01/43693 A2

(51) International Patent Classification⁷: **A61K**

[US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(21) International Application Number: **PCT/US00/34162**

(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(22) International Filing Date:
15 December 2000 (15.12.2000)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: **English**

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: **English**

Published:

(30) Priority Data:
60/172,442 17 December 1999 (17.12.1999) US

— *Without international search report and to be republished upon receipt of that report.*

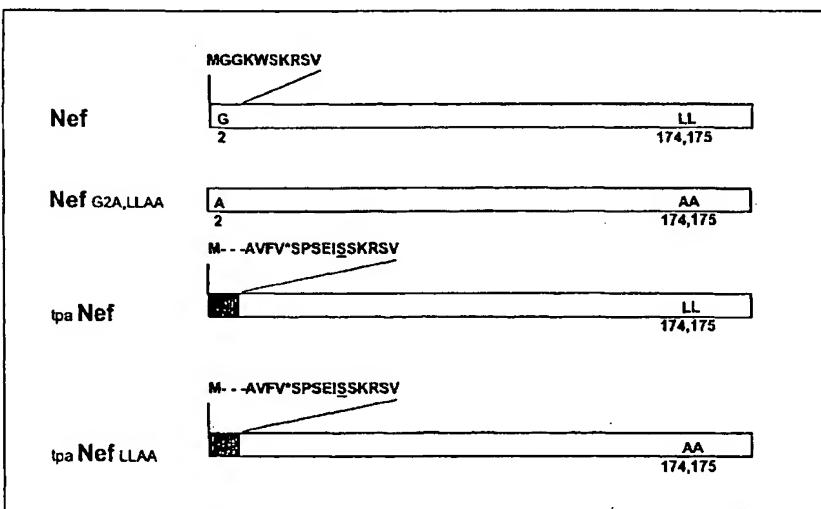
(71) Applicant (*for all designated States except US*): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). LIANG, Xiaoping [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). FU, Tong-Ming

[Continued on next page]

(54) Title: POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 NEF AND MODIFIED HIV-1 NEF



WO 01/43693 A2

(57) Abstract: Pharmaceutical compositions which comprise HIV Nef DNA vaccines are disclosed, along with the production and use of these DNA vaccines. The nef-based DNA vaccines of the invention are administered directly introduced into living vertebrate tissue, preferably humans, and express the HIV Nef protein or biologically relevant portions thereof, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The DNA molecules which comprise the open reading frame of these DNA vaccines are synthetic DNA molecules encoding codon optimized HIV-1 Nef and derivatives of optimized HIV-1 Nef, including nef modifications comprising amino terminal leader peptides, removal of the amino terminal myristylation site, and/or modification of the Nef dileucine motif. These modifications may effect wild type characteristics of Nef, such as myristylation and down regulation of host CD4.

WO 01/43693 A2



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION

5 **POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1
NEF AND MODIFIED HIV-1 NEF**

CROSS-REFERENCE TO RELATED APPLICATIONS

10 This application claims the benefit, under 35 U.S.C. §119(e), of U.S.
provisional application 60/172,442, filed December 17, 1999.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not Applicable

15

REFERENCE TO MICROFICHE APPENDIX

Not Applicable

FIELD OF THE INVENTION

20 The present invention relates to HIV Nef polynucleotide pharmaceutical
products, as well as the production and use thereof which, when directly introduced
into living vertebrate tissue, preferably a mammalian host such as a human or a
non-human mammal of commercial or domestic veterinary importance, express the
HIV Nef protein or biologically relevant portions thereof within the animal, inducing
25 a cellular immune response which specifically recognizes human immunodeficiency
virus-1 (HIV-1). The polynucleotides of the present invention are synthetic DNA
molecules encoding codon optimized HIV-1 Nef and derivatives of optimized HIV-1
Nef, including nef mutants which effect wild type characteristics of Nef, such as
myristylation and down regulation of host CD4. The polynucleotide vaccines of the
30 present invention should offer a prophylactic advantage to previously uninfected
individuals and/or provide a therapeutic effect by reducing viral load levels within an
infected individual, thus prolonging the asymptomatic phase of HIV-1 infection.

BACKGROUND OF THE INVENTION

Human Immunodeficiency Virus-1 (HIV-1) is the etiological agent of acquired human immune deficiency syndrome (AIDS) and related disorders. HIV-1 is an RNA virus of the Retroviridae family and exhibits the 5'LTR-gag-pol-env-LTR 3' organization of all retroviruses. The integrated form of HIV-1, known as the provirus, is approximately 9.8 Kb in length. Each end of the viral genome contains flanking sequences known as long terminal repeats (LTRs). The HIV genes encode at least nine proteins and are divided into three classes; the major structural proteins (Gag, Pol, and Env), the regulatory proteins (Tat and Rev); and the accessory proteins (Vpu, Vpr, Vif and Nef).

The *gag* gene encodes a 55-kilodalton (kDa) precursor protein (p55) which is expressed from the unspliced viral mRNA and is proteolytically processed by the HIV protease, a product of the *pol* gene. The mature p55 protein products are p17 (matrix), p24 (capsid), p9 (nucleocapsid) and p6.

The *pol* gene encodes proteins necessary for virus replication; a reverse transcriptase, a protease, integrase and RNase H. These viral proteins are expressed as a Gag-Pol fusion protein, a 160 kDa precursor protein which is generated via a ribosomal frame shifting. The viral encoded protease proteolytically cleaves the Pol polypeptide away from the Gag-Pol fusion and further cleaves the Pol polypeptide to the mature proteins which provide protease (Pro, P10), reverse transcriptase (RT, P50), integrase (IN, p31) and RNase H (RNase, p15) activities.

The *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity.

The *env* gene encodes the viral envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and then cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41). Gp120 and gp41 remain associated and are displayed on the viral particles and the surface of HIV-infected cells.

The *tat* gene encodes a long form and a short form of the Tat protein, a RNA binding protein which is a transcriptional transactivator essential for HIV-1 replication.

The *rev* gene encodes the 13 kDa Rev protein, a RNA binding protein. The Rev protein binds to a region of the viral RNA termed the Rev response element

(RRE). The Rev protein is promotes transfer of unspliced viral RNA from the nucleus to the cytoplasm. The Rev protein is required for HIV late gene expression and in turn, HIV replication.

- Gp120 binds to the CD4/chemokine receptor present on the surface of helper
- 5 T-lymphocytes, macrophages and other target cells in addition to other co-receptor molecules. X4 (macrophage tropic) virus show tropism for CD4/CXCR4 complexes while a R5 (T-cell line tropic) virus interacts with a CD4/CCR5 receptor complex.
- After gp120 binds to CD4, gp41 mediates the fusion event responsible for virus entry.
- 10 The virus fuses with and enters the target cell, followed by reverse transcription of its single stranded RNA genome into the double-stranded DNA via a RNA dependent DNA polymerase. The viral DNA, known as provirus, enters the cell nucleus, where the viral DNA directs the production of new viral RNA within the nucleus, expression of early and late HIV viral proteins, and subsequently the production and cellular release of new virus particles. Recent advances in the ability to detect viral load
- 15 within the host shows that the primary infection results in an extremely high generation and tissue distribution of the virus, followed by a steady state level of virus (albeit through a continual viral production and turnover during this phase), leading ultimately to another burst of virus load which leads to the onset of clinical AIDS.
- Productively infected cells have a half life of several days, whereas chronically or
- 20 latently infected cells have a 3-week half life, followed by non-productively infected cells which have a long half life (over 100 days) but do not significantly contribute to day to day viral loads seen throughout the course of disease.

- Destruction of CD4 helper T lymphocytes, which are critical to immune defense, is a major cause of the progressive immune dysfunction that is the hallmark
- 25 of HIV infection. The loss of CD4 T-cells seriously impairs the body's ability to fight most invaders, but it has a particularly severe impact on the defenses against viruses, fungi, parasites and certain bacteria, including mycobacteria.

- Effective treatment regimens for HIV-1 infected individuals have become available recently. However, these drugs will not have a significant impact on the
- 30 disease in many parts of the world and they will have a minimal impact in halting the spread of infection within the human population. As is true of many other infectious diseases, a significant epidemiologic impact on the spread of HIV-1 infection will only occur subsequent to the development and introduction of an effective vaccine. There are a number of factors that have contributed to the lack of successful vaccine

development to date. As noted above, it is now apparent that in a chronically infected person there exists constant virus production in spite of the presence of anti-HIV-1 humoral and cellular immune responses and destruction of virally infected cells. As in the case of other infectious diseases, the outcome of disease is the result of a
5 balance between the kinetics and the magnitude of the immune response and the pathogen replicative rate and accessibility to the immune response. Pre-existing immunity may be more successful with an acute infection than an evolving immune response can be with an established infection. A second factor is the considerable genetic variability of the virus. Although anti-HIV-1 antibodies exist that can
10 neutralize HIV-1 infectivity in cell culture, these antibodies are generally virus isolate-specific in their activity. It has proven impossible to define serological groupings of HIV-1 using traditional methods. Rather, the virus seems to define a serological "continuum" so that individual neutralizing antibody responses, at best, are effective against only a handful of viral variants. Given this latter observation, it
15 would be useful to identify immunogens and related delivery technologies that are likely to elicit anti-HIV-1 cellular immune responses. It is known that in order to generate CTL responses antigen must be synthesized within or introduced into cells, subsequently processed into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for
20 eventual association with major histocompatibility complex (MHC) class I proteins. CD8⁺ T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein. Activation of naive CD8⁺ T cells into activated effector or memory cells generally requires both TCR engagement of antigen as described above as well as engagement of costimulatory proteins. Optimal
25 induction of CTL responses usually requires "help" in the form of cytokines from CD4⁺ T lymphocytes which recognize antigen associated with MHC class II molecules via TCR and CD4 engagement.

As introduced above, the *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating
30 CD4 expression, disturbing T-cell activation and stimulating HIV infectivity. Zazopoulos and Haseltine (1992, *Proc. Natl. Acad. Sci.* 89: 6634-6638) disclose mutations to the HIV-1 *nef* gene which effect the rate of virus replication. The authors show that the *nef* open reading frame mutated to encode Ala-2 in place of Gly-2 inhibits myristylation of the protein and results in delayed viral replication rates

in Jurkat cells and PBMCs.

Kaminchik et al. (1991, *J. Virology* 65(2): 583-588) disclose an amino-terminal nef open reading frame mutated to encode Met-Ala-Ala in place of Met-Gly-Gly. The authors show that this mutant is deficient in myristylation.

5 Saksela et al. (1995, *EMBO J.* 14(3): 484-491) and Lee et al. (1995, *EMBO J.* 14(20): 5006-5015) show the importance of a proline rich motif in HIV-1 Nef which mediates binding to a SH3 domain of the Hck protein. The authors conclude that this motif is important in the enhancement of viral replication but not down-regulation of CD4 expression.

10 Calarota et al. (1998, *The Lancet* 351: 1320-1325) present human clinical data concerning immunization of three HIV infected individuals with a DNA plasmid expressing wild type Nef. The authors conclude that immunization with a Nef encoding DNA plasmid induced a cellular immune response in the three individuals. However, two of the three patients were on alternative therapies during the study, and
15 15 the authors conclude that the CTL response was most likely a boost to a pre-existing CTL response. In addition, the viral load increased substantially in two of the three patients during the course of the study.

20 Tobery et al. (1997, *J. Exp. Med.* 185(5): 909-920) constructed two ubiquitin-nef (Ub-nef) fusion constructs, one which encoded the Nef initiating methionine and the other with an Arg residue at the amino terminus of the Nef open reading frame. The authors state that vaccinia- or plasmid-based immunization of mice with a Ub-nef construct containing an Arg residue at the amino terminus induces a Nef-specific CTL response. The authors suggest the expressed fusion protein is more efficiently presented to the MHC class I antigen presentation pathway, resulting in an improved
25 cellular immune response.

25 Kim et al. (1997, *J. Immunol.* 158(2): 816-826) disclose that co-administration of a plasmid DNA construct expressing IL-12 with a plasmid construct expressing Nef results in an improved cellular immune response in mice when compared to inoculation with the Nef construct alone. The authors reported a reduction in the humoral response from the Nef / IL-12 co-administration as compared to
30 administration of the plasmid construct expressing Nef alone.

Moynier et al. (1998, *Vaccine* 16(16): 1523-1530) show varying humoral responses in mice immunized with a DNA plasmid encoding Nef, depending upon the presence of absence of Freund's adjuvant. No data is disclosed regarding a cellular

immune response in mice vaccinated with the aforementioned DNA construct alone.

Hanna et al. (1998, *Cell* 95:163-175) suggest that wild type Nef may play a critical role in AIDS pathogenicity.

It would be of great import in the battle against AIDS to produce a prophylactic- and/or therapeutic-based HIV vaccine which generates a strong cellular immune response against an HIV infection. The present invention addresses and meets this needs by disclosing a class of DNA vaccines based on host delivery and expression of the early HIV gene, *nef*.

10 SUMMARY OF THE INVENTION

The present invention relates to synthetic DNA molecules (also referred to herein as "polynucleotides") and associated DNA vaccines (also referred to herein as "polynucleotide vaccines") which elicit CTL responses upon administration to the host, such as a mammalian host and including primates and especially humans, as well as non-human mammals of commercial or domestic veterinary importance. The CTL-directed vaccines of the present invention should lower transmission rate to previously uninfected individuals and/or reduce levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms of HIV-1 Nef, wherein administration, intracellular delivery and expression of the HIV-1 *nef* gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Nef, codon optimized versions of HIV-1 Nef fusion proteins, and codon optimized versions of HIV-1 Nef derivatives, including but not limited to *nef* modifications involving introduction of an amino-terminal leader sequence, removal of an amino-terminal myristylation site and/or introduction of dileucine motif mutations. The Nef-based fusion and modified proteins disclosed within this specification may possess altered trafficking and/or host cell function while retaining the ability to be properly presented to the host MHC I complex and in turn elicit a host CTL and Th response.

A particular embodiment of the present invention relates to a DNA molecule encoding HIV-1 Nef from the HIV-1 jfrl isolate wherein the codons are optimized for expression in a mammalian system such as a human. The DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1, while the expressed open

reading frame is disclosed herein as SEQ ID NO:2.

In another embodiment of the present invention, a codon optimized DNA molecule encoding a protein containing the human plasminogen activator (tpa) leader peptide fused with the NH₂-terminus of the HIV-1 Nef polypeptide. The DNA

- 5 molecule which encodes this protein is disclosed herein as SEQ ID NO:3, while the expressed open reading frame is disclosed herein as SEQ ID NO:4.

In an additional embodiment, the present invention relates to a DNA molecule encoding optimized HIV-1 Nef wherein the open reading frame codes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2) and

- 10 substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175, herein described as opt nef (G2A,LLAA). The DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, while the expressed open reading frame is disclosed herein as SEQ ID NO:6.

- Another additional embodiment of the present invention relates to a DNA
15 molecule encoding optimized HIV-1 Nef wherein the amino terminal myristylation site and dileucine motif have been deleted, as well as comprising a tPA leader peptide. This DNA molecule, opt tpanef (LLAA), comprises an open reading frame which encodes a Nef protein containing a tPA leader sequence fused to amino acid residue 6-216 of HIV-1 Nef (jfrl), wherein Leu-174 and Leu-175 are substituted with Ala-174
20 and Ala-175, herein referred to as opt tpanef (LLAA) is disclosed herein as SEQ ID NO:7, while the expressed open reading frame is disclosed herein as SEQ ID NO:8.

- The present invention also relates to non-codon optimized versions of DNA molecules and associated DNA vaccines which encode the various wild type and modified forms of the HIV Nef protein disclosed herein. Partial or fully codon
25 optimized DNA vaccine expression vector constructs are preferred, but it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Nef which are shown to promote a substantial cellular immune response subsequent to host administration.

- The DNA backbone of the DNA vaccines of the present invention are
30 preferably DNA plasmid expression vectors. DNA plasmid expression vectors utilized in the present invention include but are not limited to constructs which comprise the cytomegalovirus promoter with the intron A sequence (CMV-intA) and a bovine growth hormone transcription termination sequence. In addition, the DNA plasmid vectors of the present invention preferably comprise an antibiotic resistance

marker, including but not limited to an ampicillin resistance gene, a neomycin resistance gene or any other pharmaceutically acceptable antibiotic resistance marker. In addition, an appropriate polylinker cloning site and a prokaryotic origin of replication sequence are also preferred. Specific DNA vectors of the present invention include but are not limited to V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:19.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as a prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Nef or biologically active Nef modifications or Nef-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention relate in part to codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins ligated in DNA vectors V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:19. Especially preferred DNA vaccines of the present invention include codon optimized DNA vaccine constructs V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA), as exemplified in Example Section 2.

The present invention also relates to HIV Nef polynucleotide pharmaceutical products, as well as the production and use thereof, wherein the DNA vaccines are formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention, namely by increasing a humoral response to inoculation. A preferred adjuvant is an aluminum phosphate-based adjuvant or a calcium phosphate based adjuvant,

with an aluminum phosphate adjuvant being especially preferred. Another preferred adjuvant is a non-ionic block copolymer, preferably comprising the blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. These adjuvanted forms comprising the DNA vaccines disclosed herein are useful in increasing humoral responses to DNA vaccination without imparting a negative effect on an appropriate cellular immune response.

- As used herein, a DNA vaccine or DNA polynucleotide vaccine or polynucleotide vaccine is a DNA molecule (i.e., "nucleic acid", "polynucleotide") which contains essential regulatory elements such that upon introduction into a living, vertebrate cell, it is able to direct the cellular machinery to produce translation products encoded by the respective nef genes of the present invention.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A-B show a schematic representation of DNA vaccine expression vectors V1Jns (A) and V1Jns/tpa utilized for HIV-1 nef and HIV-1 modified nef constructs.

Figure 2A-B show a nucleotide sequence comparison between wild type nef(jrf1) and codon optimized nef. The wild type nef gene from the jrf1 isolate consists of 648 nucleotides capable of encoding a 216 amino acid polypeptide. WT, wild type sequence (SEQ ID NO:9); opt, codon-optimized sequence (contained within SEQ ID NO:1). The Nef amino acid sequence is shown in one-letter code (SEQ ID NO:2).

Figure 3A-C show nucleotide sequences at junctions between nef coding sequence and plasmid backbone of nef expression vectors V1Jns/nef (Figure 3A), V1Jns/nef(G2A,LLAA) (Figure 3B), V1Jns/tpanef (Figure 3C) and V1Jns/tpanef(LLAA) (Figure 3C, also). 5' and 3' flanking sequences of codon optimized nef or codon optimized nef mutant genes are indicated by bold/italic letters; nef and nef mutant coding sequences are indicated by plain letters. Also indicated (as underlined) are the restriction endonuclease sites involved in construction of respective nef expression vectors. V1Jns/tpanef and V1Jns/tpanef(LLAA) have identical sequences at the junctions.

Figure 4 shows a schematic presentation of nef and nef derivatives. Amino acid residues involved in Nef derivatives are presented. Glycine 2 and Leucine 174

and 175 are the sites involved in myristylation and dileucine motif, respectively. For both versions of the tpanef fusion genes, the putative leader peptide cleavage sites are indicated with “*”, and a exogenous serine residue introduced during the construction of the mutants is underlined.

- 5 Figure 5 shows Western blot analysis of nef and modified nef proteins expressed in transfected 293 cells. 293 cells grown in 100 mm culture dish were transfected with respective codon optimized nef constructs. Sixty hours post transfection, supernatant and cells were collected separately and separated on 10% SDS-PAGE under reducing conditions. The proteins were transferred into a PVDF membrane and probed with a mixture of Gag mAb and Nef mAbs, both at 1:2000 dilution. The protein signals were detected with ECL. (A) cells transfected with V1Jns/gag only; (B) cells transfected with V1Jns/gag and V1Jns/nef; (C) cells transfected with V1Jns/gag and V1Jns/nef(G2A, LLAA); (D) cells transfected with V1Jns/gag and V1Jns/tpanef; (E) cells transfected with V1Jns/gag and V1Jns/tpanef(LLAA). The low case letter c and m represent medium and cellular fractions, respectively. M.W. = molecular weight marker.

- 10 Figure 6 shows an Elispot assay of cell-mediated responses to Nef peptides. Three strains of mice, Balb/c, C57BL/6 and C3H, were immunized with 50 mcg of V1Jns/nef (codon optimized) and boosted twice with a two-week interval. Two weeks following the final immunization, splenocytes were isolated and tested in an Elispot assay against respective Nef peptide pools. As a control, splenocytes were from non-immunized naive mice were tested in parallel. Nef peptide pool A consists of all 21 Nef peptides; Nef peptide pool B consists of 11 non-overlapping peptide started from residue 1; Nef peptide pool C consists of 10 non-overlapping peptides started from residue 11. SFC, INF-gamma secreting spot-forming cells.

- 15 Figure 7A-C show Nef-specific CD8 and CD4 epitope mapping. The immunization regime is as per Figure 6. Mouse splenocytes were isolated and fractionated into CD8⁺ and CD8⁻ cells using Miltenyi's magnetic cell separator. The resultant CD8⁺ and CD8⁻ cells were then tested in an Elispot assay against individual Nef peptides. SFC, INF-gamma secreting spot-forming cells. The mice strains tested are Balb/c mice (Figure 7A), C57BL/6 mice (Figure 7B), and C3H mice (Figure 7C).

20 Figure 8A-C show identification of a Nef CTL epitope. Splenocytes from nef immunized C57BL/6 mice were stimulated *in vitro* with peptide-pulsed, irradiated naïve splenocytes for 7 days. Following the *in vitro* stimulation, cells were harvested

and tested in a standard ^{51}Cr -releasing assay using peptide pulsed EL-4 cells as targets. Open symbol, specific killings of EL-4 cells without peptide; solid symbol, specific killing of EL-4 cells with peptide. Panel A - peptide Nef 51-70; Panel B - peptide Nef 60-68, Panel C - peptide Nef 58-70.

- 5 Figure 9A-B shows a comparison of the immunogenicity of codon optimized DNA vaccine vectors expressing Nef and modified forms of Nef C57BL/6 mice, five per group, were immunized with 100 mcg of the indicated nef constructs. Fourteen days following immunization, splenocytes were collected and tested against the Nef CD8 (aa58-66) and CD4 (aa81-100) peptides. Identical immunization regimens were
10 used for both experiments. In experiment 1 (Panel A), three codon optimized nef constructs were tested, namely, V1Jns/nef, V1Jns/tpanef(LLAA) and V1Jns/nef(G2A,LLAA), whereas in experiment 2 (Panel B) all four codon optimized nef constructs were tested. The data represent means plus standard deviation of 5 mice per group.

15

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to synthetic DNA molecules (also referred to herein as "nucleic acid" molecules or "polynucleotides") and associated DNA vector vaccines (also referred to herein as "polynucleotide vaccines") which elicit CTL and humoral responses upon administration to the host, including primates and especially humans. In particular, the present invention relates to DNA vector vaccines which encode various forms of HIV-1 Nef, wherein administration, intracellular delivery and expression of the HIV-1 nef gene of interest elicits a host CTL and Th response. The synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Nef, codon optimized versions of HIV-1 Nef fusion proteins, and codon optimized versions of HIV-1 Nef derivatives, including but not limited to *nef* modifications involving introduction of an amino-terminal leader sequence, removal of an amino-terminal myristylation site and/or introduction of dileucine motif mutations. In some instances the Nef-based fusion and modified proteins disclosed within this specification possess altered trafficking and/or host cell function while retaining the ability to be properly presented to the host MHC I complex. Those skilled in the art will recognize that the use of nef genes from HIV-2 strains which express Nef proteins having analogous function to HIV-1 Nef would be expected to generate immune responses analogous to those described herein for

HIV-1 constructs.

- In order to generate a CTL response, the immunogen must be synthesized within (MHC I presentation) or introduced into cells (MHC II presentation). For intracellular synthesized immunogens, the protein is expressed and then processed
- 5 into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for eventual association with major histocompatibility complex (MHC) class I proteins. CD8⁺ T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR). Activation of naive CD8⁺ T cells into activated effector or memory cells generally requires both
- 10 TCR engagement of antigen as described above as well as engagement of co-stimulatory proteins. Optimal induction of CTL responses usually requires "help" in the form of cytokines from CD4⁺ T lymphocytes which recognize antigen associated with MHC class II molecules via TCR.

The HIV-1 genome employs predominantly uncommon codons compared to

15 highly expressed human genes. Therefore, the nef open reading frame has been synthetically manipulated using optimal codons for human expression. As noted above, a preferred embodiment of the present invention relates to DNA molecules which comprise a HIV-1 nef open reading frame, whether encoding full length nef or a modification or fusion as described herein, wherein the codon usage has been

20 optimized for expression in a mammal, especially a human.

In a particular embodiment of the present invention, a DNA molecule encoding HIV-1 Nef from the HIV-1 jfrl isolate wherein the codons are optimized for expression in a mammalian system such as a human. The nucleotide sequence of the codon optimized version of HIV-1 jfrl nef gene is disclosed herein as SEQ ID NO:1,

25 as shown herein:

GATCTGCCAC CATGGGCGGC AAGTGGTCCA AGAGGTCCGT GCCCGGCTGG TCCACCGTGA
GGGAGAGGAT GAGGAGGGCC GAGCCC CGCG CCGACAGGGT GAGGAGGACC GAGCCCGCCG
CCGTGGCGT GGGCGCCGTG TCCAGGGACC TGGAGAAGCA CGGCGCCATC ACCTCCTCCA
ACACCGCCGC CACCAACGCC GACTGCGCCT GGCTGGAGGC CCAGGAGGAC GAGGAGGTGG
30 GCTTCCCCGT GAGGCCAG GTGCCCTGA GGCCCATGAC CTACAAGGGC GCCGTGGACC
TGTCCCAC TT CCTGAAGGAG AAGGGCGGCC TGGAGGGCCT GATCCACTCC CAGAAGAGGC
AGGACATCCT GGACCTGTGG GTGTACCACA CCCAGGGCTA CTTCCCGAC TGGCAGAACT
ACACCCCCGG CCCCGGCATC AGGTTCCCC TGACCTTCGG CTGGTGCTTC AAGCTGGTGC
CCGTGGAGCC CGAGAAGGTG GAGGAGGCCA ACGAGGGCGA GAACAACGTGC CTGCTGCACC

CCATGTCCCA GCACGGCATC GAGGACCCCC AGAAGGAGGT GCTGGAGTGG AGGTTCGACT
CCAAGCTGGC CTTCCACCAC GTGCCAGGG AGCTGCACCC CGAGTACTAC AAGGACTGCT
AAAGCCCGGG C (SEQ ID NO:1).

- As can be discerned from comparing native to optimized codon usage in
5 Figure 2A-B, the following codon usage for mammalian optimization is preferred:
Met (ATG), Gly (GCC), Lys (AAG), Trp (TGG), Ser (TCC), Arg (AGG), Val (GTG),
Pro (CCC), Thr (ACC), Glu (GAG); Leu (CTG), His (CAC), Ile (ATC), Asn (AAC),
Cys (TGC), Ala (GCC), Gln (CAG), Phe (TTC) and Tyr (TAC). For an additional
discussion relating to mammalian (human) codon optimization, see WO 97/31115
10 (PCT/US97/02294), which is hereby incorporated by reference.

- The open reading frame for SEQ ID NO:1 above comprises an initiating
methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides
660-662. The open reading frame of SEQ ID NO:1 provides for a 216 amino acid
15 HIV-1 Nef protein expressed through utilization of a codon optimized DNA vaccine
vector. The 216 amino acid HIV-1 Nef (jfrl) protein is disclosed herein as SEQ ID
NO:2, and as follows:

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
20 Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
25 Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
30 His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO:2).

HIV-1 Nef is a 206 amino acid cytosolic protein which associates with the
inner surface of the host cell plasma membrane through myristylation of Gly-2
(Franchini et al., 1986, *Virology* 155: 593-599). While not all possible Nef functions
have been elucidated, it has become clear that correct trafficking of Nef to the inner

plasma membrane promotes viral replication by altering the host intracellular environment to facilitate the early phase of the HIV-1 life cycle and by increasing the infectivity of progeny viral particles. In one aspect of the invention regarding codon-optimized, protein-modified polypeptides, either the DNA vaccine vector molecule or the HIV-1 nef construct is modified to contain a nucleotide sequence which encodes a heterologous leader peptide such that the amino terminal region of the expressed protein will contain the leader peptide. The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their biosynthetic pathways since their final destination, the cellular location at which they perform their function, becomes their permanent residence. Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct destinations. Defined sequence motifs exist in proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins. An effective induction of CTL responses often required sustained, high level endogenous expression of an antigen. In light of its diverse biological activities, vaccines composed of wild-type Nef could potentially have adverse effects on the host cells. As membrane-association via myristylation is an essential requirement for most of Nef's function, mutants lacking myristylation, by glycine-to-alanine change, change of the dileucine motif and/or by substitution with a tpa leader sequence as described herein, will be functionally defective, and therefore will have improved safety profile compared to wild-type Nef for use as an HIV-1 vaccine component.

In a preferred and exemplified embodiment of this portion of the invention, either the DNA vector or the HIV-1 nef nucleotide sequence is modified to include the human tissue-specific plasminogen activator (tPA) leader. As shown in Figure 1A-B for the DNA vector V1Jns, a DNA vector which may be utilized to practice the present invention may be modified by known recombinant DNA methodology to contain a leader signal peptide of interest, such that downstream

- cloning of the modified HIV-1 protein of interest results in a nucleotide sequence which encodes a modified HIV-1 tPA/Nef protein. In the alternative, as noted above, insertion of a nucleotide sequence which encodes a leader peptide may be inserted into a DNA vector housing the open reading frame for the Nef protein of interest.
- 5 Regardless of the cloning strategy, the end result is a polynucleotide vaccine which comprises vector components for effective gene expression in conjunction with nucleotide sequences which encode a modified HIV-1 Nef protein of interest, including but not limited to a HIV-1 Nef protein which contains a leader peptide. The amino acid sequence of the human tPA leader utilized herein is as follows:
- 10 MDAMKRLGCCVLLCGAVFVSPSEISS (SEQ ID NO:19).

It has been shown that myristylation of Gly-2 in conjunction with a dileucine motif in the carboxy region of the protein is essential for Nef-induced down regulation of CD4 (Aiken et al., 1994, *Cell* 76: 853-864) via endocytosis. It has also been shown that Nef expression promotes down regulation of MHC I (Schwartz et al., 15 1996, *Nature Medicine* 2(3): 338-342) via endocytosis. The present invention relates in part to DNA vaccines which encode modified Nef proteins altered in trafficking and/or functional properties. The modifications introduced into the DNA vaccines of the present invention include but are not limited to additions, deletions or substitutions to the nef open reading frame which results in the expression of a 20 modified Nef protein which includes an amino terminal leader peptide, modification or deletion of the amino terminal myristylation site, and modification or deletion of the dileucine motif within the Nef protein and which alter function within the infected host cell. Therefore, a central theme of the DNA molecules and DNA vaccines of the present invention is (1) host administration and intracellular delivery of a codon 25 optimized nef-based DNA vector vaccine; (2) expression of a modified Nef protein which is immunogenic in terms of eliciting both CTL and Th responses; and, (3) inhibiting or at least altering known early viral functions of Nef which have been shown to promote HIV-1 replication and load within an infected host.

In another preferred and exemplified embodiment of the present invention, the 30 nef coding region is altered, resulting in a DNA vaccine which expresses a modified Nef protein wherein the amino terminal Gly-2 myristylation residue is either deleted or modified to express alternate amino acid residues.

In another preferred and exemplified embodiment of the present invention, the nef coding region is altered, resulting in a DNA vaccine which expresses a modified

Nef protein wherein the dileucine motif is either deleted or modified to express alternate amino acid residues.

Therefore, the present invention relates to an isolated DNA molecule, regardless of codon usage, which expresses a wild type or modified Nef protein as described herein, including but not limited to modified Nef proteins which comprise a deletion or substitution of Gly 2, a deletion or substitution of Leu 174 and Leu 175 and/or inclusion of a leader sequence.

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

In a specific embodiment of the invention as it relates DNA vaccines encoding modified forms of HIV-1, an open reading frame which encodes a Nef protein which comprises a tPA leader sequence fused to amino acid residue 6-216 of HIV-1 Nef (jfrl) is referred to herein as opt tpanef. The nucleotide sequence comprising the open reading frame of opt tpanef is disclosed herein as SEQ ID NO:3, as shown below:

CATGGATGCA ATGAAGAGAG GGCTCTGCTG TGTGCTGCTG CTGTGTGGAG CAGTCTTCGT
TTCGCCAGC GAGATCTCCT CCAAGAGGTC CGTGCCCGGC TGTTCCACCG TGAGGGAGAG
GATGAGGAGG GCCGAGCCCCG CCGCCGACAG GGTGAGGAGG ACCGAGCCCG CCGCCGTGGG
15 CGTGGCGCC GTGTCCAGGG ACCTGGAGAA GCACGGCGCC ATCACCTCCT CCAACACCGC
CGCCACCAAAC GCGGACTGCG CCTGGCTGGA GGCCCAGGAG GACGAGGAGG TGGGCTTCCC
CGTGAGGCC CAGGTGCCCG TGAGGCCAT GACCTACAAG GGCGCCGTGG ACCTGTCCCA
CTTCCTGAAG GAGAAGGGCG GCCTGGAGGG CCTGATCCAC TCCCAGAAGA GGCAGGACAT
CCTGGACCTG TGGGTGTACC ACACCCAGGG CTACTTCCCC GACTGGCAGA ACTACACCCC
20 CGGCCCCGGC ATCAGGTTCC CCCTGACCTT CGGCTGGTGC TTCAAGCTGG TGCCCGTGGA
GCCCGAGAAG GTGGAGGAGG CCAACGAGGG CGAGAACAAAC TGCCTGCTGC ACCCCATGTC
CCAGCACGGC ATCGAGGACC CCGAGAAGGA GGTGCTGGAG TGGAGGTTCG ACTCCAAGCT
GGCCTTCCAC CACGTGGCCA GGGAGCTGCA CCCCGAGTAC TACAAGGACT GCTAAAGCC
25 (SEQ ID NO:3).

30 The open reading frame for SEQ ID NO:3 comprises an initiating methionine

residue at nucleotides 2-4 and a "TAA" stop codon from nucleotides 713-715. The open reading frame of SEQ ID NO:3 provides for a 237 amino acid HIV-1 Nef protein which comprises a tPA leader sequence fused to amino acids 6-216 of HIV-1 Nef, including the dileucine motif at amino acid residues 174 and 175. This 237
5 amino acid tPA/Nef (jfrl) fusion protein is disclosed herein as SEQ ID NO:4, and is shown as follows:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
10 Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu
Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr
Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
15 Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
Asn Cys Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
20 Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO:4).

Therefore, this exemplified Nef protein, Opt tPA-Nef, contains both a tPA leader sequence as well as deleting the myristylation site of Gly-2A DNA molecule encoding HIV-1 Nef from the HIV-1 jfrl isolate wherein the codons are optimized for expression in a mammalian system such as a human.

In another specific embodiment of the present invention, a DNA molecule is disclosed which encodes optimized HIV-1 Nef wherein the open reading frame codes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2) and substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175. This open
30 reading frame is herein described as opt nef (G2A,LLAA) and is disclosed as SEQ ID NO:5, which comprises an initiating methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides 660-662. The nucleotide sequence of this codon optimized version of HIV-1 jfrl nef gene with the above mentioned modifications is disclosed herein as SEQ ID NO:5, as follows:

GATCTGCCAC CATGGCCGGC AAGTGGTCCA AGAGGTCCGT GCCCGGCTGG TCCACCGTGA
 GGGAGAGGAT GAGGAGGGCC GAGCCCGCCG CCGACAGGGT GAGGAGGACC GAGCCCGCCG
 CCGTGGCGT GGGCGCCGTG TCCAGGGACC TGGAGAACCA CGGCCGCATC ACCTCCTCCA
 ACACCGCCGC CACCAACGCC GACTGCGCCT GGCTGGAGGC CCAGGAGGAC GAGGAGGTGG
 5 GCTTCCCCGT GAGGCCCCAG GTGCCCTGA GGCCCATGAC CTACAAGGGC GCCGTGGACC
 TGTCCCACCT CCTGAAGGAG AAGGGCGGCC TGGAGGGCCT GATCCACTCC CAGAAGAGGC
 AGGACATCCT GGACCTGTGG GTGTACCACA CCCAGGGCTA CTTCCCCGAC TGGCAGAACT
 ACACCCCCGG CCCCCGGCATC AGGTTCCCCC TGACCTTCGG CTGGTGCTTC AAGCTGGTGC
 CCGTGGAGCC CGAGAAGGTG GAGGAGGCCA ACGAGGGCGA GAACAACCTGC GCCGCCACC
 10 CCATGTCCA GCACGGCATC GAGGACCCCG AGAAGGAGGT GCTGGAGTGG AGGTTCGACT
 CCAAGCTGGC CTTCCACAC GTGCCAGGG AGCTGCACCC CGAGTACTAC AAGGACTGCT
 AAAGCCCCGGG C (SEQ ID NO:5).

The open reading frame of SEQ ID NO:5 encodes Nef (G2A,LLAA),
 disclosed herein as SEQ ID NO:6, as follows:
 15 Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
 Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
 Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
 Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
 Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
 20 Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
 Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
 Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
 Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
 Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
 25 Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Ala Ala His
 Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
 Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
 His Pro Glu Tyr Tyr Lys Asp Cys Ser (SEQ ID NO:6).

An additional embodiment of the present invention relates to another DNA
 30 molecule encoding optimized HIV-1 Nef wherein the amino terminal myristylation
 site and dileucine motif have been deleted, as well as comprising a tPA leader peptide.
 This DNA molecule, opt tpanef (LLAA) comprises an open reading frame which
 encodes a Nef protein containing a tPA leader sequence fused to amino acid residue
 6-216 of HIV-1 Nef (jfrl), wherein Leu-174 and Leu-175 are substituted with Ala-174

and Ala-175 (Ala-195 and Ala-196 in this tPA-based fusion protein). The nucleotide sequence comprising the open reading frame of opt tpanef (LLAA) is disclosed herein as SEQ ID NO:7, as shown below:

CATGGATGCA ATGAAGAGAG GGCTCTGCTG TGTGCTGCTG CTGTGTGGAG CAGTCCTCGT
 5 TTCGCCAGC GAGATCTCCT CCAAGAGGTC CGTGCCCGGC TGGTCCACCG TGAGGGAGAG
 GATGAGGAGG GCCGAGCCCG CGGCCGACAG GGTGAGGAGG ACCGAGCCCG CGGCCGTGGG
 CGTGGCGCC GTGTCCAGGG ACCTGGAGAA GCACGGGCC ATCACCTCCT CCAACACCGC
 CGCCACCAAC CGCGACTGCG CCTGGCTGGA GGCCCAGGAG GACGAGGAGG TGGGCTTCCC
 CGTGAGGCC CAGGTGCCCG TGAGGCCAT GACCTACAAG GGCGCCGTGG ACCTGTCCCA
 10 CTTCCCTGAAG GAGAAGGGCG GCCTGGAGGG CCTGATCCAC TCCCAGAAGA GGCAGGACAT
 CCTGGACCTG TGGGTGTACC ACACCCAGGG CTACTTCCCC GACTGGCAGA ACTACACCCC
 CGGCCCGGC ATCAGGTTCC CCCTGACCTT CGGCTGGTGC TTCAAGCTGG TGCCCGTGG
 GCCCGAGAAG GTGGAGGAGG CCAACGAGGG CGAGAACAAAC TGCGCCGCC ACCCCATGTC
 CCAGCACGGC ATCGAGGACC CCGAGAAGGA GGTGCTGGAG TGGAGGTTCG ACTCCAAGCT
 15 GGCCCTCCAC CACGTGGCCA GGGAGCTGCA CCCCGAGTAC TACAAGGACT GCTAAAGCCC
 (SEQ ID NO:7).

The open reading frame of SEQ ID NO:7 encoding tPA-Nef (LLAA), disclosed herein as SEQ ID NO:8, is as follows:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 20 Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro
 Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala
 Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val
 Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala
 Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu
 25 Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr
 Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
 Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
 Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
 Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
 30 Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn
 Asn Cys Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
 Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
 Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO:8).

The present invention also relates in part to any DNA molecule, regardless of

- codon usage, which expresses a wild type or modified Nef protein as described herein, including but not limited to modified Nef proteins which comprise a deletion or substitution of Gly 2, a deletion or substitution of Leu 174 and Leu 175 and/or inclusion of a leader sequence. Therefore, partial or fully codon optimized DNA
- 5 vaccine expression vector constructs are preferred since such constructs should result in increased host expression. However, it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Nef which are shown to promote a substantial cellular immune response subsequent to host administration.
- 10 The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors are well known in the art and the present DNA vector vaccines may be comprised of any such expression backbone which contains at least a promoter for RNA polymerase transcription, and a transcriptional terminator 3' to the HIV nef coding sequence. In
- 15 one preferred embodiment, the promoter is the Rous sarcoma virus (RSV) long terminal repeat (LTR) which is a strong transcriptional promoter. A more preferred promoter is the cytomegalovirus promoter with the intron A sequence (CMV-intA). A preferred transcriptional terminator is the bovine growth hormone terminator. In addition, to assist in large scale preparation of an HIV nef DNA vector vaccine, an antibiotic resistance marker is also preferably included in the expression vector.
- 20 Ampicillin resistance genes, neomycin resistance genes or any other pharmaceutically acceptable antibiotic resistance marker may be used. In a preferred embodiment of this invention, the antibiotic resistance gene encodes a gene product for neomycin resistance. Further, to aid in the high level production of the pharmaceutical by
- 25 fermentation in prokaryotic organisms, it is advantageous for the vector to contain an origin of replication and be of high copy number. Any of a number of commercially available prokaryotic cloning vectors provide these benefits. In a preferred embodiment of this invention, these functionalities are provided by the commercially available vectors known as pUC. It is desirable to remove non-essential DNA
- 30 sequences. Thus, the lacZ and lacI coding sequences of pUC are removed in one embodiment of the invention.

DNA expression vectors exemplified herein are also disclosed in PCT International Application No. PCT/US94/02751, International Publication No. WO 94/21797, hereby incorporated by reference. A first DNA expression vector

is the expression vector pnRSV, wherein the rous sarcoma virus (RSV) long terminal repeat (LTR) is used as the promoter. A second embodiment relates to plasmid V1, a mutated pBR322 vector into which the CMV promoter and the BGH transcriptional terminator is cloned. Another embodiment regarding DNA vector backbones relates 5 to plasmid V1J. Plasmid V1J is derived from plasmid V1 and removes promoter and transcription termination elements in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. Therefore, V1J also contains the CMVintA promoter and (BGH) transcription 10 termination elements which control the expression of the HIV nef-based genes disclosed herein. The backbone of V1J is provided by pUC18. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. The entire *lac* operon was removed and the remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA 15 polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element. In another DNA expression vector, the ampicillin resistance gene is removed from V1J and replaced with a neomycin resistance gene, to generate V1Jneo. A DNA expression vector specifically exemplified herein is V1Jns, which is the same as V1J except that a unique Sfi1 restriction site has been engineered into the single Kpn1 site at position 2114 of V1J-neo. The incidence of Sfi1 sites in human 20 genomic DNA is very low (approximately 1 site per 100,000 bases). Thus, this vector allows careful monitoring for expression vector integration into host DNA, simply by Sfi1 digestion of extracted genomic DNA. Another DNA expression vector for use as the backbone to the HIV-1 nef-based DNA vaccines of the present invention is V1R. In this vector, as much non-essential DNA as possible is "trimmed" from the vector to 25 produce a highly compact vector. This vector is a derivative of V1Jns. This vector allows larger inserts to be used, with less concern that undesirable sequences are encoded and optimizes uptake by cells when the construct encoding specific influenza virus genes is introduced into surrounding tissue.

It will be evident upon review of the teaching within this specification that 30 numerous vector/Nef antigen constructs may be generated. While the exemplified constructs (V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA) are preferred, any number of vector/Nef antigen combinations are within the scope of the present invention, especially wild type or modified Nef proteins which comprise a deletion or substitution of Gly 2, a deletion of substitution of Leu 174 and Leu 175

and/or inclusion of a leader sequence. Therefore, the present invention especially relates to DNA vaccines and a pharmaceutically active vaccine composition which contains this DNA vector vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention include but in no way are limited to codon optimized DNA molecules encoding HIV-1 Nef of biologically active Nef modifications or Nef-containing fusion proteins ligated in DNA vectors V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:19. Especially preferred DNA vaccines of the present invention include as V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA), as exemplified in Example Section 2.

The DNA vector vaccines of the present invention may be formulated in any pharmaceutically effective formulation for host administration. Any such formulation may be, for example, a saline solution such as phosphate buffered saline (PBS). It will be useful to utilize pharmaceutically acceptable formulations which also provide long-term stability of the DNA vector vaccines of the present invention. During storage as a pharmaceutical entity, DNA plasmid vaccines undergo a physiochemical change in which the supercoiled plasmid converts to the open circular and linear form. A variety of storage conditions (low pH, high temperature, low ionic strength) can accelerate this process. Therefore, the removal and/or chelation of trace metal ions (with succinic or malic acid, or with chelators containing multiple phosphate ligands) from the DNA plasmid solution, from the formulation buffers or from the vials and closures, stabilizes the DNA plasmid from this degradation pathway during storage. In addition, inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, are useful to prevent damage of the DNA plasmid from free radical production that may still occur, even in apparently demetalated solutions. Furthermore, the buffer type, pH, salt concentration, light exposure, as well as the

type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine. Therefore, formulations that will provide the highest stability of the DNA vaccine will be one that includes a demetalated solution containing a buffer (phosphate or bicarbonate) with a pH in the 5 range of 7-8, a salt (NaCl, KCl or LiCl) in the range of 100-200 mM, a metal ion chelator (e.g., EDTA, diethylenetriaminepenta-acetic acid (DTPA), malate, inositol hexaphosphate, tripolyphosphate or polyphosphoric acid), a non-reducing free radical scavenger (e.g. ethanol, glycerol, methionine or dimethyl sulfoxide) and the highest appropriate DNA concentration in a sterile glass vial, packaged to protect the highly 10 purified, nuclease free DNA from light. A particularly preferred formulation which will enhance long term stability of the DNA vector vaccines of the present invention would comprise a Tris-HCl buffer at a pH from about 8.0 to about 9.0; ethanol or glycerol at about 3% w/v; EDTA or DTPA in a concentration range up to about 5 mM; and NaCl at a concentration from about 50 mM to about 500 mM. The use of 15 such stabilized DNA vector vaccines and various alternatives to this preferred formulation range is described in detail in PCT International Application No. PCT/US97/06655, PCT International Publication No. WO 97/40839, which is hereby incorporated by reference.

The DNA vector vaccines of the present invention may, in addition to 20 generating a strong CTL-based immune response, provide for a measurable humoral response subsequent immunization. This response may occur with or without the addition of adjuvant to the respective vaccine formulation. To this end, the DNA vector vaccines of the present invention may also be formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA 25 polynucleotide vaccines of the present invention. A number of these adjuvants are known in the art and are available for use in a DNA vaccine, including but not limited to particle bombardment using DNA-coated gold beads, co-administration of DNA vaccines with plasmid DNA expressing cytokines, chemokines, or costimulatory molecules, formulation of DNA with cationic lipids or with 30 experimental adjuvants such as saponin, monophosphoryl lipid A or other compounds which increase immunogenicity of the DNA vaccine. One preferred adjuvant for use in the DNA vector vaccines of the present invention are one or more forms of an aluminum phosphate-based adjuvant. Aluminum phosphate is known in the art for use with live, killed or subunit vaccines, but is only recently

disclosed as a useful adjuvant in DNA vaccine formulations. The artisan may alter the ratio of DNA to aluminum phosphate to provide for an optimal immune response. In addition, the aluminum phosphate-based adjuvant possesses a molar PO₄/Al ratio of approximately 0.9, and may again be altered by the skilled artisan

5 to provide for an optimal immune response. An additional mineral-based adjuvant may be generated from one or more forms of a calcium phosphate. These mineral-based adjuvants are useful in increasing humoral responses to DNA vaccination without imparting a negative effect on an appropriate cellular immune response. Complete guidance for use of these mineral-based compounds for use

10 as DNA vaccines adjuvants are disclosed in PCT International Application No. PCT/US98/02414, PCT International Publication No. WO 98/35562, which are hereby incorporated by reference in their entirety. Another preferred adjuvant is a non-ionic block copolymer which shows adjuvant activity with DNA vaccines. The basic structure comprises blocks of polyoxyethylene (POE) and

15 polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. (1998, *Critical Reviews in Therapeutic Drug Carrier Systems* 15(2): 89-142) review a class of non-ionic block copolymers which show adjuvant activity. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. *id.*, disclose

20 that certain POE-POP-POE block copolymers may be useful as adjuvants to an influenza protein-based vaccine, namely higher molecular weight POE-POP-POE block copolymers containing a central POP block having a molecular weight of over about 9000 daltons to about 20,000 daltons and flanking POE blocks which comprise up to about 20% of the total molecular weight of the copolymer (see also

25 U.S. Reissue Patent No. 36,665, U.S. Patent No. 5,567,859, U.S. Patent No. 5,691,387, U.S. Patent No. 5,696,298 and U.S. Patent No. 5,990,241, all issued to Emanuele, et al., regarding these POE-POP-POE block copolymers).

WO 96/04932 further discloses higher molecular weight POE/POP block copolymers which have surfactant characteristics and show biological efficacy as

30 vaccine adjuvants. The above cited references within this paragraph are hereby incorporated by reference in their entirety. It is therefore within the purview of the skilled artisan to utilize available adjuvants which may increase the immune response of the polynucleotide vaccines of the present invention in comparison to administration of a non-adjuvanted polynucleotide vaccine.

The DNA vector vaccines of the present invention are administered to the host by any means known in the art, such as enteral and parenteral routes. These routes of delivery include but are not limited to intramuscular injection, intraperitoneal injection, intravenous injection, inhalation or intranasal delivery, oral delivery, sublingual administration, subcutaneous administration, transdermal administration, transcutaneous administration, percutaneous administration or any form of particle bombardment, such as a biolistic device such as a "gene gun" or by any available needle-free injection device. The preferred methods of delivery of the HIV-1 Nef-based DNA vaccines disclosed herein are intramuscular injection and needle-free injection. An especially preferred method is intramuscular delivery.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1 μ g to 15 greater than about 20 mg, and preferably in doses from about 1 mg to about 5 mg is administered directly into muscle tissue. As noted above, subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, inhalation and oral delivery are also contemplated. It is also contemplated that booster vaccinations are to be 20 provided in a fashion which optimizes the overall immune response to the Nef-based DNA vector vaccines of the present invention.

The aforementioned polynucleotides, when directly introduced into a vertebrate *in vivo*, express the respective HIV-1 Nef protein within the animal and in turn induce a cytotoxic T lymphocyte (CTL) response within the host to the expressed 25 Nef antigen. To this end, the present invention also relates to methods of using the HIV-1 Nef-based polynucleotide vaccines of the present invention to provide effective immunoprophylaxis, to prevent establishment of an HIV-1 infection following exposure to this virus, or as a post-HIV infection therapeutic vaccine to mitigate the acute HIV-1 infection so as to result in the establishment of a lower virus 30 load with beneficial long term consequences. As noted above, the present invention contemplates a method of administration or use of the DNA nef-based vaccines of the present invention using any of the known routes of introducing polynucleotides into living tissue to induce expression of proteins.

Therefore, the present invention provides for methods of using a DNA nef-

based vaccine utilizing the various parameters disclosed herein as well as any additional parameters known in the art, which, upon introduction into mammalian tissue induces *in vivo*, intracellular expression of these DNA nef-based vaccines. This intracellular expression of the Nef-based immunogen induces a CTL and humoral 5 response which provides a substantial level of protection against an existing HIV-1 infection or provides a substantial level of protection against a future infection in a presently uninfected host.

The following examples are provided to illustrate the present invention without, however, limiting the same hereto.

10

EXAMPLE 1

Vaccine Vectors

V1 – Vaccine vector V1 was constructed from pCMVIE-AKI-DHFR (Whang et al., 1987, *J. Virol.* 61: 1796). The AKI and DHFR genes were removed by cutting 15 the vector with EcoRI and self-ligating. This vector does not contain intron A in the CMV promoter, so it was added as a PCR fragment that had a deleted internal SacI site [at 1855 as numbered in Chapman, et al., (1991, *Nuc. Acids Res.* 19: 3979)]. The template used for the PCR reactions was pCMVintA-Lux, made by ligating the HindIII and NheI fragment from pCMV6a120 (see Chapman et al., *ibid.*), which 20 includes hCMV-IE1 enhancer/promoter and intron A, into the HindIII and XbaI sites of pBL3 to generate pCMVIntBL. The 1881 base pair luciferase gene fragment (HindIII-SmaI Klenow filled-in) from RSV-Lux (de Wet et al., 1987, *Mol. Cell Biol.* 7: 725) was ligated into the Sall site of pCMVIntBL, which was Klenow filled-in and phosphatase treated. The primers that spanned intron A are: 5' primer: 25 5'-CTATATAAGCAGAGCTCGTTAG-3' (SEQ ID NO:10); 3' primer: 5'-GTAGCAAAGATCTAAGGACGGTGACTGCAG-3' (SEQ ID NO:11). The primers used to remove the SacI site are: sense primer, 5'-GTATGTGTCTG AAAATGAGC GTGGAGATTGGGCTCGCAC-3' (SEQ ID NO:12) and the antisense primer, 5'-GTGCGAGCCCAATCTCCACGCTCATTTCAGAC 30 ACATA-3' (SEQ ID NO:13). The PCR fragment was cut with Sac I and Bgl II and inserted into the vector which had been cut with the same enzymes.

V1J – Vaccine vector V1J was generated to remove the promoter and transcription termination elements from vector V1 in order to place them within a more defined context, create a more compact vector, and to improve plasmid

purification yields. V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment. pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-
5 characterized by sequence and function, and is of small size. The entire *lac* operon was removed from this vector by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase treated with calf intestinal alkaline phosphatase, and
10 ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of
15 two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in *E. coli* and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher
20 expression of heterologous genes compared with V1. The nucleotide sequence of V1J is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGCTCT GTAAGCGGAT GCGGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCAGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
25 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCCGCGT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
30 TGCCCACATTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA

CTCCGCCCCA TTGACGCAA TGCGCGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
 AGCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACCGCTGTT TTGACCTCCA
 TAGAAGACAC CGGGACCAGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCA CCCCCTTGGC
 5 TTCTTATGCA TGCTATACTG TTTTGGCTT GGGGTCTATA CACCCCCGCT TCCTCATGTT
 ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC
 CTATTGGTGA CGATACTTC CATTACTAAT CCATAACATG GCTCTTGCC ACAACTCTCT
 TTATTGGCTA TATGCCAATA CACTGTCCCT CAGAGACTGA CACGGACTCT GTATTTTAC
 AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC
 10 CCCAGTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTCCGG
 ACATGGGCTC TTCTCCGGTA GCGCGGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
 CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCTAACAA GTGGAGGCCA GACTTAGGCA
 CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
 TGAAAATGAG CTCGGGGAGC GGGCTTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC
 15 GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC
 CGTTGGGTG CTGTTAACGG TGGAGGGAGC TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
 GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTCCA TGGGTCTTTT
 CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGTTGCC
 CCTCCCCGT GCCTTCTTG ACCCTGGAAG GTGCCACTCC CACTGTCCCTT TCCTAATAAA
 20 ATGAGGAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTG TATTCTGGGG GGTGGGGTGG
 GGCAAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAC
 AGGCACATCC CCTTCTCTGT GACACACCCT GTCCACGCC CTGGTTCTTA GTTCCAGGCC
 CACTCATAGG ACACTCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
 25 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG
 TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
 CGCTCGGTGCG TTGGCTGCG GCGAGCGGT ACAAAGCGGT AATACGGTTA
 TCCACAGAAT CAGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCC GAAAAGGCC
 30 AGGAACCGTA AAAAGGCCGC GTTGTGGCG TTTTCCATA GGCTCCGCC CCCTGACGAG
 CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCCT GCCGCTTACC
 GGATACCTGT CCGCCTTCT CCCTCGGGGA AGCGTGGGCC TTTCTCAATG CTCACGCTGT
 AGGTATCTCA GTTGGGTGTA GGTCGTTGCG TCCAAGCTGG GCTGTGTGCA CGAACCCCCC

GTTCACCCCG ACCGCTGCCG CTTATCCGGT AACTATCGTC TTGAGTCAA CCCGGTAAGA
 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
 GGCAGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA
 5 TCCGGCAAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTG TTTGCAAGCA GCAGATTACG
 CGCAGAAAAA AAGGATCTCA AGAACATCCT TTGATCTTT CTACGGGGTC TGACGCTCAG
 TGGAAACGAAA ACTCACGTTA AGGGATTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
 TAGATCCTTT TAAATTAAAA ATGAAGTTT AAATCAATCT AAAGTATATA TGAGTAAACT
 TGGCTGACA GTTACCAATG CTTAACAGT GAGGCACCTA TCTCAGCGAT CTGTCTATT
 10 CGTTCATCCA TAGTTGCCGT ACTCCCCGTC GTGTAGATAA CTACGATAACG GGAGGGCTTA
 CCATCTGGCC CCAGTGCTGC AATGATAACCG CGAGACCCAC GCTCACCGGC TCCAGATTAA
 TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGCCTGC AACTTTATCC
 GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT
 AGTTTGCAGCA ACGTTGTGC CATTGCTACA GGCAATCGTGG TGTCAAGCCTC GTCTGTTGGT
 15 ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG
 TGCAAAAAG CGGTTAGCTC CTTCGGTCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA
 GTGTTATCAC TCATGGTTAT GGCAAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA
 AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG
 CGACCGAGTT GCTCTTGGCC GGCGTCAATA CGGGATAATA CGCGGCCACA TAGCAGAACT
 20 TTAAAAGTGC TCATCATTGG AAAACGTTCT TCAGGGCGAA AACTCTCAAG GATCTTACCG
 CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTT
 ACTTTCACCA GCGTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA
 ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC
 ATTATCAGG GTTATTGTCT CATGAGCGGA TACATATTG AATGTATTAA GAAAAATAAA
 25 CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC CTGACGTCTA AGAAACCATT
 ATTATCATGA CATTAAACCTA TAAAAATAGG CGTATCACGA GGCCCTTCG TC (SEQ ID
 NO:14).

V1Jneo – Construction of vaccine vector V1Jneo expression vector involved removal of the amp^r gene and insertion of the kan^r gene (neomycin phosphotransferase). The amp^r gene from the pUC backbone of V1J was removed by digestion with SspI and Eam1105I restriction enzymes. The remaining plasmid was purified by agarose gel electrophoresis, blunt-ended with T4 DNA polymerase, and then treated with calf intestinal alkaline phosphatase. The commercially available kan^r gene, derived from transposon 903 and contained within the pUC4K plasmid,

was excised using the PstI restriction enzyme, purified by agarose gel electrophoresis, and blunt-ended with T4 DNA polymerase. This fragment was ligated with the V1J backbone and plasmids with the kan^r gene in either orientation were derived which were designated as V1Jneo #'s 1 and 3. Each of these plasmids was confirmed by 5 restriction enzyme digestion analysis, DNA sequencing of the junction regions, and was shown to produce similar quantities of plasmid as V1J. Expression of heterologous gene products was also comparable to V1J for these V1Jneo vectors. V1Jneo#3, referred to as V1Jneo hereafter, was selected which contains the kan^r gene in the same orientation as the amp^r gene in V1J as the expression construct and 10 provides resistance to neomycin, kanamycin and G418. The nucleotide sequence of V1Jneo is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGCTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
15 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
20 CATACTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
TGGCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
25 CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAT GTCGTAACAA
CTCCGCCCA TTGACGCAA TGCGGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTA GTGAACCGTC AGATGCCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGAA CGGTGCATTG GAACGCGGAT
TCCCCGTGCC AAGAGTGACG TAAAGTACCGC CTATAGAGTC TATAAGGCCA CCCCCTTGGC
30 TTCTTATGCA TGCTATACTG TTTTGCGCTT GGGGTCTATA CACCCCGCT TCCTCATGTT
ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC
CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTGCC ACAACTCTCT
TTATTGGCTA TATGCCAATA CACTGTCCCT CAGAGACTGA CACGACTCT GTATTTTAC
AGGATGGGT CTCATTTATT ATTTACAAAT TCACATATAAC AACACCACCG TCCCCAGTGC

CCGCAGTTT TATTAACAT AACGTGGAT CTCCACCGA ATCTCGGT ACGTGTCCGG
ACATGGGCTC TTCTCCGGTA GCGCGGGAGC TTCTACATCC GAGCCCTGCT CCCATGCC
CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAAC A GTGGAGGCCA GACTTAGGCA
CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
5 TGAAAATGAG CTCGGGGAGC GGGCTTGAC CGCTGACGCC TTTGGAAGAC TTAAGGCAGC
GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACCTCC
CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCC
GCGGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTCCA TGGGTCTTT
CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGTTGCC
10 CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCCTT TCCTAATAAA
ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC TATTCTGGGG GGTGGGGTGG
GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAAGAAC
AGGCACATCC CCTTCTCTGT GACACACCC GTCCACGCC CTGGTTCTTA GTTCCAGCCC
15 CACTCATAGG ACACCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG
TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
CCCTCGGTG TTGGCTGCG GCGAGCGTA TCAGCTCACT CAAAGCGGT AATACGGTA
20 TCCACAGAAT CAGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC
AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTCCATA GGCTCCGCC CCCTGACGAG
CATCACAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCCT GCCGCTTACC
GGATACCTGT CCGCTTCTCC CCGCTGGGA AGCGTGGCGC TTTCTCAATG CTCACGCTGT
25 AGGTATCTCA GTTCGGTGT A GTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCC
GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA
CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
GGCGGTGCTA CAGAGTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTGTA
30 TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTG TTTGCAAGCA CGAGATTACG
CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTT CTACGGGGTC TGACGCTCAG
TGGAACGAAA ACTCACGTTA AGGGATTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
TAGATCCTTT TAAATTAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT

CGTTCATCCA TAGTTGCCTG ACTCCGGGGG GGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
 AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA
 GCCACGGTTG ATGAGAGCTT TGTTGTAGGT GGACCAAGTTG GTGATTTGTA ACTTTTGCTT
 TGCCACGGAA CGGTCTGCGT TGTCGGGAAG ATGCGTGATC TGATCCTCA ACTCAGCAA
 5 AGTCGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
 TACAACCAAT TAACCAATTG TGATTAGAAA AACTCATCGA GCATCAAATG AACTGCAAT
 TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTCTG TAATGAAGGA
 GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
 ACTCGTCAA CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAAATAAG GTTATCAAGT
 10 GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT
 TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
 AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACCGCATE GCTGTTAAAA
 GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACAGTGCAG CGCATCAACA
 ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTT CCCGGGGATC
 15 GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTGGAAAGA
 GGCATAAATT CCGTCAGCCA GTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
 CTACCTTGC CATGTTTCAG AAACAACCTCT GGCGCATCGG GCTTCCCATA CAATCGATAG
 ATTGTCGCAC CTGATTGCC GACATTATCG CGAGCCCATT TATAACCCATA TAAATCAGCA
 TCCATGTTGG AATTAAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
 20 ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTA TTGTTCATGA TGATATATT
 TTATCTTGTG CAATGTAACA TCAGAGATT TGAGACACAA CGTGGCTTTC CCCCCCCCC
 CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
 TAGAAAAATA AACAAATAGG GGTCAGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTT
 25 CGTC (SEQ ID NO:15).

V1Jns - The expression vector V1Jns was generated by adding an SfiI site to V1Jneo to facilitate integration studies. A commercially available 13 base pair SfiI linker (New England BioLabs) was added at the KpnI site within the BGH sequence of the vector. V1Jneo was linearized with KpnI, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt SfiI linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with SfiI) was comparable to expression of the same genes in V1Jneo (with KpnI).

The nucleotide sequence of V1Jns is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
5 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCCGCGT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGACTATT TACGGTAAAC
10 TGCCCACCTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCAATTATGC CCAGTACATG ACCTTATGGG ACTTTCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
CGTCAATGGG AGTTTGTGTT GGCAACAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
15 CTCCGCCCA TTGACGCAAA TGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
ACCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACCGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGACTC TATAGGCACA CCCCTTGGC
TCTTATGCAT GCTATACTGT TTTGGCTTG GGGCCTATAC ACCCCCCGTT CCTTATGCTA
20 TAGGTGATGG TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
TATTGGTGAC GATACTTCC ATTACTAAC TATAACATGG CTCTTGCCA CAACTATCTC
TATTGGCTAT ATGCCAATAC TCTGCTCTTC AGAGACTGAC ACGGACTCTG TATTTTTACA
GGATGGGGTC CCATTTATTA TTACAAATT CACATATACA ACAACGCCGT CCCCCGTGCC
CGCAGTTTT ATTAAACATA GCGTGGGATC TCCACCGCAA TCTCGGTAC GTGTTCCGG
25 CATGGGCTCT TCTCCGGTAG CGGGGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
AGCGGCTCAT GGTGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG ACTTAGGCAC
AGCACAAATGC CCACCACCCAC CAGTGTGCCG CACAAGCCCG TGGCGGTAGG GTATGTGTCT
GAAAATGAGC GTGGAGATTG GGCTCGCAGC GCTGACGCCAG ATGGAAGACT TAAGGCAGCG
GCAGAAGAAG ATGCAGGCCAG CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACCTCCC
30 GTTGCAGGTGC TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCCG
CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTCCAT GGGTCTTTTC
TGCAGTCACC GTCCTTAGAT CTGCTGTGCC TTCTAGTTGC CAGCCATCTG TTGTTGCC
CTCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC ACTGTCCTTT CCTAATAAAA
TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCATTCT ATTCTGGGGG GTGGGGTGGG

GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG CATGCTGGGG ATGCCGGTGGG
CTCTATGGCC GCTCGGCC GGTGCTGAAG AATTGACCCG GTTCCTCCTG GGCCAGAAAG
AAGCAGGCAC ATCCCCTCT CTGTGACACA CCCTGTCCAC GCCCCTGGTT CTTAGTTCCA
GCCCCACTCA TAGGACACTC ATAGCTCAGG AGGGCTCCGC CTTCAATCCC ACCCGCTAAA
5 GTACTTGGAG CGGTCTCTCC CTCCCTCATC AGCCCACCAA ACCAAACCTA GCCTCCAAGA
GTGGGAAGAA ATTAAAGCAA GATAGGCTAT TAAGTGCAGA GGGAGAGAAA ATGCCTCCAA
CATGTGAGGA AGTAATGAGA GAAATCATAG AATTTCTTCC GCTTCCTCGC TCACTGACTC
GCTGCGCTCG GTCGTTGGC TGCGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG
GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA
10 GGCCAGGAAC CGTAAAAGG CCGCGTTGCT GGCGTTTTC CATAGGCTCC GCCCCCTGA
CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG
ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCCTCT CCTGTTCCGA CCCTGCCGCT
TACCGGATAC CTGTCGCCCT TTCTCCCTTC GGGAAGCGTG GCGCTTCTC ATAGCTCAGC
CTGTAGGTAT CTCAGTCGG TGTAGGTCGT TCGCTCCAAG CTGGCTGTG TGCACGAACC
15 CCCCGTTCAAG CCCGACCGCCT GCGCCTTATC CGGTAACATCGT CCAACCCGGT
AAGACACGAC TTATGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA
TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC
AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC
TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTGCA AGCAGCAGAT
20 TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC
TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT
CACCTAGATC CTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA
AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT
ATTTCGTTCA TCCATAGTTG CCTGACTCGG GGGGGGGGGG CGCTGAGGTC TGCTCGTGA
25 AGAAGGTGTT GCTGACTCAT ACCAGGCCTG AATCGCCCCA TCATCCAGCC AGAAAGTGAG
GGAGGCCACGG TTGATGAGAG CTTTGTGTA GGTGGACCAAG TTGGTGATTT TGAACCTTTG
CTTTGCCACG GAACGGCTCG CGTTGTGGG AAGATGCGTG ATCTGATCCT TCAACTCAGC
AAAAGTTCGA TTTATTCAAC AAAGCCGCCG TCCCCTCAAG TCAGCGTAAT GCTCTGCCAG
TGTACAAACC AATTAACCA TTCTGATTAG AAAAACTCAT CGACCATCAA ATGAAACTGC
30 AATTTATTCA TATCAGGATT ATCAATACCA TATTTTGAA AAAGCCGTTT CTGTAATGAA
GGAGAAAAGT CACCGAGGCA GTTCCATAGG ATGGCAAGAT CCTGGTATCG GTCTGCGATT
CCGACTCGTC CAACATCAAT ACAACCTATT AATTTCCCT CGTCAAAAT AAGGTTATCA
AGTGAGAAAT CACCATGAGT GACCGACTGAA TCCGGTGAGA ATGGCAAAAG CTTATGCATT
TCTTCCAGA CTTGTTCAAC AGGCCAGCCA TTACGCTCGT CATCAAAATC ACTCGCATCA

ACCAAACCGT TATTCATTG TGATTGCGCC TGAGCGAGAC GAAATACGCG ATCGCTGTTA
 AAAGGACAAT TACAAACAGG AATCGAATGC AACCGGGCGCA GGAACACTGC CAGCGCATCA
 ACAATATTTT CACCTGAATC AGGATATTCT TCTAATACCT GGAATGCTGT TTTCCCGGGG
 ATCGCAGTGG TGAGTAACCA TGCATCATCA GGAGTACGGA TAAAATGCTT GATGGTCGGA
 5 AGAGGCATAA ATTCCGTCAG CCAGTTAGT CTGACCACCT CATCTGTAAC ATCATTGGCA
 ACGCTACCTT TGCCATGTTT CAGAAACAAC TCTGGCGCAT CGGGCTTCCC ATACAATCGA
 TAGATTGTCG CACCTGATTG CCCGACATTA TCGCGAGCCC ATTTATACCC ATATAAATCA
 GCATCCATGT TGGAATTAA TCGCGGCCTC GAGCAAGACG TTTCCCGTTG AATATGGCTC
 10 ATAACACCCC TTGTATTACT GTTATGTAA GCAGACAGT TTATTGTTCA TGATGATATA
 TTTTATCTT GTGCAATGTA ACATCAGAGA TTTTGAGACA CAACGTGGCT TTCCCCCCCC
 CCCCATTTA GAAGCATTAA TCAGGGTTAT TGTCTCATGA GCGGATAACAT ATTTGAATGT
 ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC
 GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCC
 15 TTCGTC (SEQ ID NO:16).

15 The underlined nucleotides of SEQ ID NO:16 represent the SfiI site introduced into the Kpn 1 site of V1Jneo.

V1Jns-tPA – The vaccine vector V1Jns-tPA was constructed in order to fuse an heterologous leader peptide sequence to the nef DNA constructs of the present invention. More specifically, the vaccine vector V1Jns was modified to include the 20 human tissue-specific plasminogen activator (tPA) leader. As an exemplification, but by no means a limitation of generating a nef DNA construct comprising an amino-terminal leader sequence, plasmid V1Jneo was modified to include the human tissue-specific plasminogen activator (tPA) leader. Two synthetic complementary oligomers were annealed and then ligated into V1Jneo which had been BglII digested. The 25 sense and antisense oligomers were 5' GATCACCATGGATGCAATGAAGAGAG GGCTCTGCTGTGCTGCTGTGGAGCAGTCTCGTTGCCAG CGA-3' (SEQ ID NO:17); and, 5'-GATCTCGCTGGCGAACGAAGACTGC TCCACACAGCAGCAGCACACAGCAGAGCCCTCTTCATTGCATCCAT GGT-3' (SEQ ID NO:18). The Kozak sequence is underlined in the sense oligomer.
 30 These oligomers have overhanging bases compatible for ligation to BglII-cleaved sequences. After ligation the upstream BglII site is destroyed while the downstream BglII is retained for subsequent ligations. Both the junction sites as well as the entire tPA leader sequence were verified by DNA sequencing. Additionally, in order to conform with V1Jns (=V1Jneo with an SfiI site), an SfiI restriction site was placed at

the KpnI site within the BGH terminator region of V1Jneo-tPA by blunting the KpnI site with T4 DNA polymerase followed by ligation with an SfiI linker (catalogue #1138, New England Biolabs), resulting in V1Jns-tPA. This modification was verified by restriction digestion and agarose gel electrophoresis.

- 5 The V1Jns-tpa vector nucleotide sequence is as follows:

TCGC CGC GTTT CGGT GATGAC GGT GAAA ACC TCT GAC ACAT GC AG CT CCG G GAG AC CG GT CA
CAG CTT GT CT GTA AG CGG AT GCC GGG AG CA GACA AG CCG TC AG GG CG CG TC AG CG GGG TG
TT GG CG GG TG TC GGG GCT GG CTT AACT ATG CGG CAT CAGA GC AG ATT GT TA CT GAG AGT G C
ACC AT ATG CG GT GT GAA ATA CGG CAC AG AT GCG TA AGG AG AAA AT ACC GC AT CAG ATT GG
10 CT AT TGG CCA TT GCATA CGT TGT AT CCATA TCATA AT ATG TAC AT TTA TATA TT GG CT CAT G
TCC AAC ATTA CCG CCAT GT TT GAC AT TG ATT ATT GACT AGT TAT TA AT AGT AAT CA ATT AC
GGG GT CAT TA GTT CAT AG CC CAT AT ATG GA GTT CC CG GTT AC AT AACT TA CGG TAA AT GG
CCC GCG CT GG C TG ACC GCG CC ACG ACC CCG CCC AT TG AC G TCA AT AT GA CG TAT GTT CC
CAT AG TA AC G CCA AT AG GG A CTT CC AT TG ACG TCA AT GG GT GG AGT ATT TAC CG GT AA AC
15 TG CCC ACT TG GC AGT AC ATC AAG TGT AT CA TAT GCA AAG T ACG CCCC CT A TT GAC GT CAA
TG AC GG TAA TG GCG CC CT GG CATT AT GC CC AGT AC ATG AC CT TAT GGG ACT TT CCT AC
TT GG CAG TAC AT CT AC GT AT TAG TCA TGC T ATT ACC ATG GT GAT GCG GT TT GG CAG TA
CAT CA AT GGG CGT GG AT AGC GG TT GACT C AC GGG GAT TT CCA AGT CT CC ACC CATT GA
CGT CA AT GGG AG TT GT TT GG CAC CAAA TCA AC GGG AC TT TCCA AA AT GTC GTA AC AA
20 CT CC GCG CC CA TT GAC GCA AA TG GCG GT AG GCG TGT AC GG TG GG AGG TCT AT AT AAG CAG
AG CT CG TT TA GT GAC CCG TC AG AT CG CT G GAG AC GCG CAT CC AC GCT GT TT GAC CT CC A
TAG AAG AC AC CGG GAC CG AT CC AG CCG CT CG CG GCG CCG GAA CG GT GC AT TG GAA CG CG AT
TCC CGT GCG AAG AGT GAC G TA AGT ACC GCG CT AT AGA CT C TAT AGG CACA CCC TT TT GGC
TCT TAT GC AT GCT AT ACT GT TTT GG CT TG GGG C CT AT AC ACC CCG CT TT CCT TAT GCT A
25 TAG GT GAT GG TAT AG CT TAG CCT AT AG GT TG GGG T TAT TG ACC AT TAT TG ACC ACT CCCC
TAT TGG TGA C GATA CT TT CC ATT ACT AAT C CATA AC AT GG CT CT TG GCC CAA CT AT CTC
TAT TGG CT AT ATGCC AAT AC TCT GT C CT AG ACT GAC ACG GACT CTG TAT TTT TACA
GG AT GGG GTC CC AT TT AT TA TT TACA AA AT CAC AT ACA AC AC GCG CT CCCC CGT GCG
CG CAG TT TT ATT AA AC AT A GCG TGG GAT C TCC AC GCG GAA TCT CGG TAC GT GT CC GGA
30 CAT GGG CT CT TCT CC GGT AG CGG CGG AG CT TCC AC AT CG AG CC CT GG TC CC AT GCG CT CC
AG CG GCT CAT GG TCG CT CG G CAG CT C CT TG CTC CT A AC AG TGG AGG CC AG ACT TAGG CAC
AG CACA AT GC CC ACC ACC AC CAG TGT GCG CACA AGG CCG TGG CGG TAG G TAT GT GT CT
GAA AA AT GAG C GT GG AG AT TG GG CT CG CAC G GCT GAC GCG AG AT GGA AG ACT TA AGG CAG CG
GC AG AAG AAG AT GC CAG GCG AG CT GAG TT GT TGT AT AAG AGT CAG A GG TA ACT CCCC

GTTGGGGTGC TGTAAACGGT GGAGGGCAGT CTAGTCTGAG CACTACTCGT TGCTGCCGCG
 CGGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTCAT GGGTCTTTTC
TGCAGTCACC GTCCTTAGAT CACCATGGAT GCAATGAAGA GAGGGCTCTG CTGTGTGCTG
CTGCTGTGTG GAGCAGTCTT CGTTTCGCCC AGCGAGATCT GCTGTGCCTT CTAGTTGCCA
 5 GCCATCTGTT GTTGCCCCCT CCCCCGTGCC TTCCCTGACC CTGGAAGGTG CCACTCCCAC
TGCTCTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT CTGAGTAGGT GTCATTCTAT
TCTGGGGGT GGGGTGGGGC AGGACAGCAA GGGGGAGGAT TGGGAAGACA ATAGCAGGCA
TGCTGGGGAT GCGGTGGGCT CTATGGCCGC TGCGGCCAGG TGCTGAAGAA TTGACCCGGT
TCCTCCTGGG CCAGAAAGAA GCAGGCACAT CCCCTCTCT GTGACACACC CTGTCCACGC
 10 CCCTGGTTCT TAGTTCCAGC CCCACTCATA GGACACTCAT AGCTCAGGAG GGCTCCGCCT
TCAATCCCAC CCGCTAAAGT ACTTGGAGCG GTCTCTCCCT CCCTCATCAG CCCACCAAAC
CAAACCTAGC CTCCAAGAGT GGGAAAGAAAT TAAAGCAAGA TAGGCTATTA AGTGCAGAGG
GAGAGAAAAT GCCTCCAACA TGTGAGGAAG TAATGAGAGA AATCATAGAA TTTCTTCCGC
TTCTCGCTC ACTGACTCGC TGCGCTCGT CGTTCGGCTG CGCGGAGCGG TATCAGCTCA
 15 CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG
AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTCCA
TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA
CCCGACAGGA CTATAAGAT ACCAGGCCTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC
TGTTCCGACC CTGCCGCTTA CCGGATAACCT GTCCGCCTT CTCCCCTCGG GAAGCGTGGC
 20 GCTTCTCAT AGCTCACGCT GTAGGTATCT CAGTCGGTG TAGTCGTTG GCTCCAAGCT
GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATATCG
TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG
GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA
CGGCTACACT AGAAGAACAG TATTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG
 25 AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACACC GCTGGTAGCG GTGGTTTTT
TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT
TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT
CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC
 30 TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCGGGG GGGGGGGCG
CTGAGGTCTG CCTCGTGAAG AAGGTGTTGC TGACTCATAC CAGGCCCTGAA TCGCCCCATC
ATCCAGCCAG AAAGTGAGGG AGCCACGGTT GATGAGAGCT TTGTTGTAGG TGGACCAGTT
GGTGTATTTG AACTTTGCT TTGCCACGGA ACGGTCTGCG TTGTCGGGAA GATGCGTGAT
CTGATCCTTC AACTCAGCAA AAGTTCGATT TATTCAACAA AGCCGCCGTC CCGTCAAGTC

AGCGTAATGC TCTGCCAGTG TTACAACCAA TTAACCAATT CTGATTAGAA AACTCATCG
 AGCATCAAAT GAAACTGCAA TTTATTCTATA TCAGGATTAT CAATACCATA TTTTGAAAA
 AGCCGTTTCT GTAAATGAAGG AGAAAACCTCA CCGAGGCAGT TCCATAGGAT GGCAAGATCC
 TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC AACCTATTAA TTTCCCCCTCG
 5 TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA CGACTGAATC CGGTGAGAAT
 GGCAAAAGCT TATGCATTTC TTTCCAGACT TGTTCAACAG GCCAGCCATT ACGCTCGTCA
 TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCTG ATTGCGCCTG AGCGAGACGA
 AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA CCGGCGCAGG
 AACACTGCCA GCGCATCAAC AATATTTCA CCTGAATCAG GATATTCTTC TAATACCTGG
 10 AATGCTGTTT TCCCAGGGAT CGCAGTGGTG AGTAACCAGT CATCATCAGG AGTACGGATA
 AAATGCTTGA TGGTCGAAG AGGCATAAAAT TCCGTCAGCC AGTTTAGTCT GACCATCTCA
 TCTGTAACAT CATTGGCAAC GCTACCTTG CCATGTTCA GAAACAACTC TGGCGCATCG
 GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC GCGAGCCCAT
 TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAAC GCGGCCTCGA GCAAGACGTT
 15 TCCCAGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT TTATGTAAGC AGACAGTTTT
 ATTGTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT TTGAGACACA
 ACGTGGCTTT CCCCCCCCCC CCATTATTGA AGCATTATC AGGGTTATTG TCTCATGAGC
 GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC
 CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
 20 AGGCGTATCA CGAGGCCCTT TCGTC (SEQ ID NO:9).

The underlined nucleotides of SEQ ID NO:9 represent the Sfi1 site introduced into the Kpn 1 site of V1Jneo while the underlined/italicized nucleotides represent the human tPA leader sequence.

V1R – Vaccine vector V1R was constructed to obtain a minimum-sized
 25 vaccine vector without unneeded DNA sequences, which still retained the overall
 optimized heterologous gene expression characteristics and high plasmid yields that
 V1J and V1Jns afford. It was determined that (1) regions within the pUC backbone
 comprising the *E. coli* origin of replication could be removed without affecting
 plasmid yield from bacteria; (2) the 3'-region of the *kan^r* gene following the
 30 kanamycin open reading frame could be removed if a bacterial terminator was
 inserted in its place; and, (3) ~300 bp from the 3'- half of the BGH terminator could
 be removed without affecting its regulatory function (following the original KpnI
 restriction enzyme site within the BGH element). V1R was constructed by using PCR
 to synthesize three segments of DNA from V1Jns representing the CMVintA

promoter/BGH terminator, origin of replication, and kanamycin resistance elements, respectively. Restriction enzymes unique for each segment were added to each segment end using the PCR oligomers: SspI and XhoI for CMVintA/BGH; EcoRV and BamHI for the *kan* r gene; and, BclI and SalI for the *ori* r. These enzyme sites
5 were chosen because they allow directional ligation of each of the PCR-derived DNA segments with subsequent loss of each site: EcoRV and SspI leave blunt-ended DNAs which are compatible for ligation while BamHI and BclI leave complementary overhangs as do SalI and XhoI. After obtaining these segments by PCR each segment was digested with the appropriate restriction enzymes indicated above and then
10 ligated together in a single reaction mixture containing all three DNA segments. The 5'-end of the *ori* r was designed to include the T2 rho independent terminator sequence that is normally found in this region so that it could provide termination information for the kanamycin resistance gene. The ligated product was confirmed by restriction enzyme digestion (>8 enzymes) as well as by DNA sequencing of the
15 ligation junctions. DNA plasmid yields and heterologous expression using viral genes within V1R appear similar to V1Jns. The net reduction in vector size achieved was 1346 bp (V1Jns = 4.86 kb; V1R = 3.52 kb). PCR oligomer sequences used to synthesize V1R (restriction enzyme sites are underlined and identified in brackets following sequence) are as follows: (1) 5'-GGTACAAATATTGGCTATTGGC
20 CATTGCATACG-3' (SEQ ID NO:20) [SspI]; (2) 5'-CCACATCTCGAGGAA CCGGGTCAATTCTTCAGCACC-3' (SEQ ID NO:21) [XhoI] (for CMVintA/BGH segment); (3) 5'-GGTACAGATATCGGAAAGCCACGTTGTG TCTAAAATC-3' (SEQ ID NO:22) [EcoRV]; (4) 5'-CACATGGATCCGTAATGCTCTGCCAGTGT TACAACC-3' (SEQ ID NO:23) [BamHI], (for kanamycin resistance gene segment)
25 (5) 5'-GGTACCATG ATCACGTAGAAAAGATCAAAGGATCTTCTTG-3' (SEQ ID NO:24) [BclI]; (6) 5'-CCACATGTCGACCCGTAAAAAGGCCCGTTGCTGG-3' (SEQ ID NO:25): [SalI], (for *E. coli* origin of replication).

The nucleotide sequence of vector V1R is as follows:

TCGGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
30 CAGCTTGCT GTAAGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTAAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC

GGGGTCACTTA GTTCATAGCC CATATATGGA GTTCCCGCGTT ACATAACTTA CGGTAAATGG
 CCCGCCTGGC TGACCGCCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
 CATACTAACG CCAATAGGGG CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
 TGCCCACCTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 5 TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCCTAC
 TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
 CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
 CTCCGCCCA TTGACGCAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
 10 AGCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACCGCTGTT TTGACCTCCA
 TAGAACGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGAA CGGTGCATTG GAACGCGGAT
 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCA CCCCCTTGGC
 TTCTTATGCA TGCTATACTG TTTTGCGCTT GGGGTCTATA CACCCCGCT TCCTCATGTT
 ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC
 15 CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTGCC ACAACTCTCT
 TTATTGGCTA TATGCCAATA CACTGTCCCT CAGAGACTGA CACGACTCT GTATTTTAC
 AGGATGGGAT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC
 CCGCAGTTTT TATTAAACAT AACGTGGGAT CTCCACGGGA ATCTCGGGTA CGTGTCCGG
 ACATGGGCTC TTCTCCGGTA CGGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCCTC
 20 CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCTAACAA GTGGAGGCCA GACTTAGGCA
 CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
 TGAAAATGAG CTCGGGGAGC GGGCTTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC
 GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC
 CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
 25 GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCTTTCCA TGGGTCTTTT
 CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTCTGCC
 CCTCCCCCGT GCCTTCTTG ACCCTGGAAG GTGCCACTCC CACTGTCCCTT TCCTAATAAA
 ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCAATTC TATTCTGGGG GGTGGGGTGG
 GGCAGCACAG CAAGGGGGAG GATGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 30 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCTGGGCC AGAAAGAAC
 AGGCACATCC CCTTCTCTGT GACACACCCT GTCCACGCC CTGGTTCTTA GTTCCAGCCC
 CACTCATAGG ACACTCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG

TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
CGCTCGGTGCG TTCGGCTGCG GCGAGCGGT A TCAGCTCACT CAAAGGC GGTT AATACGGTTA
TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC
AGGAACCGTA AAAAGGCCG G GTTGCTGGCG TTTTCGCATA GGCTCCGCC CCCTGACGAG
5 CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
CAGGC GTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCCT CCCGCTTACCG
GGATACCTGT CCGCCTTCT CCCTCGGG A GCGTGGCGC TTTCTCAATG CTCACGCTGT
AGGTATCTCA GTTCGGTGTA GGTGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC
GTTCA GCGCC ACCGCTGCGC CTTATCCGGT AACTATCGC TTGAGTCAA CCCGGTAAGA
10 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAA CAGGA TTAGCAGAGC GAGGTATGTA
GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA
TCCGGCAAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTG TTTGCAAGCA GCAGATTACG
CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTT CTACGGGTC TGACGCTCAG
15 TGGAACGAAA ACTCACGTTA AGGGATTTG GTCATGAGAT TATCAAAAG GATCTTCACC
TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
TGGTCTGACA GTTACCAATG CTTAATCACT GAGGCACCTA TCTCAGCGAT CTGTCTATT
CGTTCATCCA TAGTTGCCTG ACTCCGGGGGG GGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGG
20 GCCACGGTTG ATGAGAGCTT TGTGTTAGGT GGACCAGTTG GTGATTTGA ACTTTTGCTT
TGCCACGGAA CGGTCTGCGT TGTGGGAAG ATGCGTGATC TGATCCTTCA ACTCAGCAAA
AGTTGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG AAACTGCAAT
TTATTCAAT CAGGATTATC AATACCATAT TTTTGAAAAA GCGTTTCTG TAATGAAGGA
25 GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCGCTGT CAAAATAAG GTTATCAAGT
GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTCT
TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACCGGATC GCTGTTAAA
30 GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA
ATATTTTCACT CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTT CCCGGGGATC
GCAGTGGTGA GTAACCAGTC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTGGAAGA
GGCATAAAATT CCGTCAGCCA GTTGTGCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
CTACCTTGC CATGTTTCAG AAACAACCTCT GGCGCATCGG GCTTCCCATA CAATCGATAG

ATTGTCGCAC CTGATTGCC GACATTATCG CGAGCCCATT TATAACCCATA TAAATCAGCA
 TCCATGTTGG AATTAAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
 ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTA TTGTTCATGA TGATATATTT
 TTATCTTGTG CAATGTAACA TCAGAGATT TGAGACACAA CGTGGCTTTC CCCCCCCCCC
 5 CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
 TAGAAAAATA AACAAATAGG GGTCGCCGC ACATTTCCCC GAAAAGTGC ACCTGACGTC
 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTT
 CGTC (SEQ ID NO:26).

10

EXAMPLE 2

Codon Optimized HIV-1 Nef and HIV-1 Nef Derivatives as DNA Vector Vaccines

HIV-1 Nef Vaccine Vectors - Codon optimized nef gene coding for wt Nef protein of HIV-1 jrf1 isolate was assembled from complementary, overlapping synthetic oligonucleotides by polymerase chain reaction (PCR). The PCR primers used were designed in such that a BglII site was included in the extension of 5' primer and an SrfI site and a BglII site in the extension of 3' primer. The PCR product was digested with BglII and cloned into BglII site of a human cytomeglovirus early promoter-based expression vector, V1Jns (Figure 1A). The proper orientation of nef fragment in the context of the expression cassette was determined by asymmetric restriction mapping. The resultant plasmid is V1Jns/nef. The 5' and 3' nucleotide sequence junctions of codon optimized V1Jns/nef are shown in Figure 3A.

The mutant nef (G2A,LLAA) was also made from synthetic oligonucleotides. To assist in cloning, a PstI site and an SrfI site were included in the extensions of 5' and 3' PCR primers, respectively. The PCR product was digested with PstI and SrfI, and cloned into the PstI and SrfI sites of V1Jns/nef, replacing the original nef with nef(G2A,LLAA) fragment. This resulted in V1Jns/nef(G2A,LLAA). The 5' and 3' nucleotide sequence junctions of codon optimized V1Jns/nef (G2A,LLAA) are shown in Figure 3B.

To construct the expression vector containing human tissue plasminogen activator leader peptide and the nef fusion gene, i.e., V1Jns/tPAnef, a truncated nef gene fragment, lacking the coding sequence for the five amino terminal residues, was first amplified by PCR using V1Jns/nef as template. Both 5' and 3' PCR primers used in this reaction contained a BglII extension. The PCR amplified fragment was then digested with BglII and cloned into BglII site of the expression vector, V1Jns/tpa

(Figure 1B). The ligation of the 3' end of tpa leader peptide coding sequence to the 5' end of the nef PCR product restored the BglII site and yielded an in-frame fusion of the two genes. The 5' and 3' nucleotide sequence junctions of codon optimized V1Jns/tPAnef are shown in Figure 3C.

5 Construction of V1Jns/tpanef(LLAA) was carried out by replacing the Bsu36-SacII fragment of V1Jns/tpanef, which contains the 3' half of the nef gene and part of the vector backbone, with the Bsu36-SacII fragment from V1Jns/nef(G2A,LLAA). The 5' and 3' nucleotide sequence junctions of codon optimized V1Jns/tpanef (LLAA) are shown in Figure 3C.

10 All the nef constructs were verified by sequencing. The amino acid junctions of these constructs is shown schematically in Figure 4.

Transfection and protein expression - 293 cells (adenovirus transformed human embryonic kidney cell line 293) grown at approximately 30% confluence in minimum essential medium (MEM; GIBCO, Grand Island, MD) supplemented with
15 10% fetal bovine serum (FBS; GIBCO) in a 100 mm culture dish, were transfected with 4 ug gag expression vector, V1Jns/gag, or a mixture of 4 ug gag expression vector and 4 ug nef expression vector by Lipofectin following manufacturer's protocol (GIBCO). Twelve hours post-transfection, cells were washed once with 10 ml of serum-free medium, Opti-MEM I (GIBCO) and replenished with 5 ml of Opti-MEM.
20 Following an additional 60 hr incubation, culture supernatants and cells were collected separately and used for Western blot analysis.

Western blot analysis - Fifty microliter of samples were separated on a 10% SDS-polyacrylamide gel (SDS-PAGE) under reducing conditions. The proteins were blotted onto a piece of PVDF membrane, and reacted to a mixture of gag mAb (#18;
25 Intracel, Cambridge, MA) and Nef mAbs (aa64-68, aa195-201; Advanced Biotechnologies, Columbia, MD), both at 1:2000 dilution, and horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (Zymed, San Francisco, CA). The protein bands were visualized by ECL Western blotting detection reagents, according to the manufacturer's protocol (Amersham, Arlington Heights, IL).

30 Enzyme-linked immunosorbent assay (ELISA) - 96-well Immulon II, round-bottom plates were coated with 50 ul of Nef protein at the concentration of 2ug/ml in bicarbonate buffer, pH 9.8., per well at 4°C overnight. Plates were washed three times with PBS containing 0.05% Tween-20 (PBST), and blocked with 5% skim milk in PBST (milk-PBST) at 24°C for 2 hr, and then incubated with serial dilutions of

testing samples in milk-PBST at 24°C for 2 hr. Plates were washed with PBST three times, and added with 50 ul of HRP-conjugated goat anti-mouse IgG (Zymed) per well and incubated at 24°C for 1 hr. This was followed by three washes, and the addition of 100 ul of 1 mg/ml ABTS [(2,2'-amino-di-(3-ethylbenzthiozoline sulfonate)] (KPL, Gaithersburg, MD) per well. After 1 hr at 24°C, plates were read at a wavelength of 405nm using an ELISA plate reader.

- 5 *Enzyme-linked spot assay (Elispot)* - Nitrocellulose membrane-backed 96 well plates (MSHA plates; Millipore, Bedford, MA) were coated with 50 ul of rat anti-mouse IFN-gamma mAb, capture antibody, (R4-6A2; PharMingen, San Diego, CA) at 10 a concentration of 5ug/ml in PBS per well at 4°C overnight. Plates were washed three times with PBST and blocked with 10% FBS in RPMI-1640 (FBS-RPMI) at 37°C in a CO₂ incubator for 2 to 4 hrs. Splenocytes were suspended in RPMI-1640 with 10% FBS at 4 x 10⁶ cells per ml. 100 ul cells were added to each well and plates were incubated at 37°C for 20 hrs. Each sample was tested in triplicate wells. After 15 incubation, plates were rinsed briefly with distilled water and washed three times with PBST. Fifty ul of biotinylated rat anti-mouse IFN-γ mAb, detecting antibody (XMG1.2; PharMingen), diluted in 1% BSA in PBST at a concentration of 2 ug/ml was then added to each well. Plates were incubated at 24°C for 2 hr, followed by washes with PBST. Fifty ul of streptavidin-conjugated alkaline phosphatase (KPL) at 20 a dilution of 1:1000 in FBS-RPMI was added to each well. The plates were incubated at 24C for an additional one hr. Following extensive wash with BPST, 100ul BCIT/NBT substrate (KPL) was added for 15 min, and color reaction was stopped by washing the plate with tap water. Plates were air-dried and spots were counted using a dissection microscope.
- 25 *Cytotoxic T cell (CTL) assay* - Splenocytes from immunized mouse were co-cultured with syngenic peptide-pulsed, irradiated naive splenocytes for 7 days. EL-4 cells were incubated at 37°C for 1 hr with or without 20ug/ml of a designated peptide in the presence of sodium 51Cr-chromate and used as target cells. For the assay, 10⁴ target cells were added to a 96-well plate along with different numbers of splenocytes 30 cells. Plates were incubated at 37°C for 4 hr. After incubation, supernatants were collected and counted in a Wallac gamma-counter. Specific lysis was calculated as ([experimental release - spontaneous release]/maximum release- spontaneous release]) x 100%. Spontaneous release was determined by incubating target cells in

medium alone, and maximum release was determined by incubating target cells in 2.5% TritonX-100. The assay was performed with triplicate samples.

Animal experiments - Female mice (Charles River Laboratories, Wilmington, MA), 6 to 10 weeks old, were injected in quadriceps with 100 ul of DNA in PBS.

- 5 Two weeks after immunization, spleens from individual mice were collected and used for CTL and Elispot assays.

Results (DNA Vector Vaccine Construction) - The exemplified Nef protein sequence is based on HIV-1 clade B jrfl isolate. A codon-optimized nef gene was chosen for vaccine construction and for use as the parental gene for other exemplified constructs. Figure 2A-B show the comparison of coding sequence of wt nef(jrfl) and the codon optimized nef(jrfl). Two forms of myristylation site mutations were constructed; one contains a.Gly2Ala change and the other a human tissue plasminogen activator (tpa) leader sequence was fused to sixth residue, Ser, of Nef (tpanef). The dileucine motif mutation was made by introducing both Leu174Ala and 10 Leu175Ala changes. Figure 4 shows the schematic depiction of the Nef and Nef mutants. For *in vitro* expression and *in vivo* immunogenicity studies, the nef genes were cloned into expression vector, V1Jns. The resultant plasmids containing wt nef, tpanef, tpanef with dileucine motif mutation, and nef mutant with the Gly2Ala 15 myristylation site and dileucine motif mutations were named as V1Jns/nef, V1Jns/tpanef, V1Jns/tpanef(LLAA) and V1Jns/(G2A,LLAA), respectively.

Results - Expression and Western blotting analysis - To evaluate the expression of the codon optimized nef constructs, adenovirus-transformed human kidney 293 cells were cotransfected with individual nef plasmids and a gag expression vector, V1Jns/gag. 72 hours post transfection, cells and medium were collected 20 separately and analyzed by Western blotting, using both Nef- and Gag-specific mAbs. The results are shown in Figure 5. Cells transfected with V1Jns/gag only revealed a single distinct band of approximately 55 Kd, whereas the cells cotransfected with gag and nef plasmids revealed, in addition to the 55 Kd band, a major 30 Kd band and several minor bands. This pattern is consistent with that the 55 Kd species represents 25 Gag polypeptide and the 30 Kd and other minor species are the Nef-related products. Therefore, all the nef constructs were expressed in the transfected cells. When measured against the relatively constant Gag signal as a reference, four nef genes seem to be expressed at different levels, with the following descending order, tpanef, nef, tpanef(LLAA) and nef(G2A, LLAA). With the exception of nef(G2A,LLAA), 30

products of nef, tpanef, tpanef(LLAA) could be detected in both cellular and medium fractions.

- Mapping of Nef-specific CD8 and CD4 epitopes in mice* – There was no information available with respect to the properties of Nef(jrfl) in eliciting cell-mediated immune responses in mice. Therefore, to characterize immunogenicity of Nef and Nef mutants exemplified herein, CD8 and CD4 epitopes were mapped. An overlapping set of overlapping nef peptides that encompass the entire 216 aa Nef polypeptide were generated. A total 21 peptides were made, which include twenty 20mers and one 16mer. Three strains of mice, Balb/c, C3H and C57BL/6, were immunized with plasmid V1Jns/Nef; splenocytes from immunized and naive mice were isolated and assessed for Nef specific INF-gamma secreting cells (SFC) by the Elispot assay. Figure 6 shows where Elispot assays were performed against separate pools of the Nef peptides. All three strains of immunized mice responded to the Nef plasmid immunization; each developed positive Nef peptide-specific INF- γ SFCs.
- Based on this, further studies were carried out with fractionated CD8 and CD4 cells against individual peptides. The results are shown in Figure 7A-C. In Balb/c mice (Figure 7A), four Nef peptides, namely, aa11-30, aa61-80, aa191-210 and aa200-216, were found to be able to induce significant numbers of CD4 SFCs. In C57BL/6 mice (Figure 7B), only one peptide, ie., aa81-100, elicited significant numbers of CD4 SFCs. Compared to Balb/c and C57BL/6 mice, C3H mice (Figure 7C) showed no dominant CD4 SFC responses with particular peptides; instead, there were modest number of SFCs in response to an array of peptides, including aa21-40, aa31-50, aa121-140 aa131-150, aa181-200 and aa191-210. With respect to CD8 cells, significant SFC responses were detected with a single peptide, ie., aa51-70, in C57BL/6 mice only.

The results from Elispot assay suggested that Nef peptide aa51-70 contained an H-2b restricted CD8 cell epitope. In order to ascertain whether this CD8 epitope also represents the cytotoxic T cell (CTL) epitope, a conventional CTL assay was carried out. The peptide aa51-70 (Figure 8A) induced low level of specific killings only. Peptides longer than 9 amino acids of a typical CTL epitope often have lower binding affinity to MHC class I molecule. It was contemplated that the low specific killings observed with peptide aa51-70 could be potentially resulted from the low binding affinity of this 20 amino acid peptide. Therefore, two shortened peptides, namely, aa60-68 and aa58-70, were synthesized and tested in CTL assays. While the

peptide aa60-68 failed to elicit any specific killings (Figure 8B), the peptide aa58-70 exhibited a drastic increase of specific killing as compared to its longer counterpart, peptide aa61-80 (Figure 8C). For example, the percentage of specific killings induced by peptide aa58-70 at an effector/target ratio of 5 to 1 was comparable to that induced
 5 by peptide aa51-80 at an effector/target ratio of 45. Thus, between peptide aa58-70 and peptide aa51-70, the former was almost ten-fold more effective in terms of inducing Nef-specific killing. The results from CTL assay therefore confirmed that the CD8 epitope detected by the Elispot assay was indeed a CTL epitope. To further map the minimum amino acid sequence for the Nef CTL epitope, additional 5
 10 peptides were synthesized and analyzed by Elispot assay, which mapped the CTL epitope to Nef aa58-66, as shown in Table 1.

TABLE 1

Nef peptides**	INF-γ SFC*/10 ⁶ splenocytes
Nef58-70	TAATNADCAWLEA
Nef59-69	AATNADCAWLE
Nef58-68	TAATNADCAWL
Nef58-67	TAATNADCAW
Nef58-66	TAATNADCA
Medium	1

* Average of duplicate samples.
 15 ** Amino acid sequence of all peptides contained within SEQ ID NO:2.

Results (Evaluation of Immunogenicity of nef Mutants in Mice) - Having identified H-2b restricted CTL and CD4 cell epitopes, the immunogenicity of the different codon optimized nef constructs in C57BL/6 mice was examined. This was 5 performed in two separate experiments with identical immunization regimens. The first experiment involved nef, tpanef(LLAA) and nef(G2A,LLAA) and the second experiment involved nef, tpanef, tpanef(LLAA) and nef(G2A,LLAA). Mice were immunized with plasmids containing these respective codon optimized nef genes. Two weeks post immunization, splenocytes from individual mice were isolated and 10 analyzed by Elispot assay for Nef-specific CD8 and CD4 IFN-gamma SFCs using Nef peptide aa58-66 and aa81-100, respectively. The results are shown in Figure 9A-B. In the experiment 1 (Figure 9A), among the three groups tested, the mice receiving the codon optimized tpanef(LLAA) construct developed the highest CD8 and CD4 cell responses; comparing between tpanef(LLAA) and the nef, the former elicited 15 about 40-fold higher CD8 SFCs and 10-fold higher CD4 SFCs. In contrast to tpanef(LLAA), nef(G2A,LLAA) mutant was poorly immunogenic; mice receiving this mutant had barely detectable CD8 and CD4 SFCS, under conditions tested. Similar response profiles between the three mutants were also observed in the 20 experiment 2 (Figure 9B), except that the overall CD8 response of mice receiving tpanef(LLAA) was approximately 10-folder higher in experiment 2 than that observed in experiment 1. The tPAnef mutant showed comparable responses as that of tpanef(LLAA). The results therefore showed that both codon optimized tpanef and tpanef(LLAA) had significantly enhanced immunogenicity.

Results (Evaluation of Immunogenicity of nef Mutants in Rhesus Monkeys) - 25 Monkeys were immunized with 5 mg of indicated codon optimized plasmids at week 0, 4, and 8. Four weeks after each immunization , peripheral blood mononuclear cells were collected and tested for Nef-specific INF-gamma secreting cells as described for the mice studies in this Example section. The results are shown in Table 2. As with the mouse study, tpanef(LLAA) shows significantly enhanced 30 immunogenicity when compared to tPAnef.

TABLE 2

Vaccine	Animal No.	Nef specific INF-gamma secreting cells/million PBMC					
		Week 0		Week 4		Week 8	
		Medium	nef	Medium	nef	Medium	nef
VIJns-TpaNef (LLAA)	1	74	39	30	208	6	148
	2	1	3	28	45	13	44
	3	5	5	14	45	11	11
VIJns-nef	1	0	1	24	33	16	43
	2	28	9	31	35	13	34
	3	1	0	16	31	18	38
Control	1	1	3	16	33	16	16
						18	13

- 5 Monkeys were immunized with 5 mg of indicated plasmids at week 0, 4 and 8.
 Four weeks after each immunization, peripheral blood mononuclear cells were collected and tested for the Nef-specific IFN-gamma secreting cells.
- 10 A codon-optimized nef gene coding for HIV-1 jrf1 isolate Nef polypeptide was synthesized. The resultant synthetic nef gene was well expressed in the *in vitro* transfected cells. Using this synthetic gene as parental molecule, nef mutants involving myristylation site and dileucine motif mutations were constructed. Two forms of myristylation site mutation were made, one involving a single Gly2Ala change and the other by fusing human plasminogen activator(tpa) leader peptide with the N-terminus of Nef polypeptide. The dileucine motif mutation was generated by 15 Leu174Ala and Leu175Ala changes. The resultant nef constructs were named as nef, tpanef, tpanef(LLAA) and nef(G2A,LLAA). The addition of tpa leader peptide sequence resulted in significantly increased expression of the nef gene *in vitro*; in contrast, either Gly2Ala mutation or dileucine mutation reduced the nef gene

expression. In an effort to characterize immunogenicity of nef and nef mutants, experiments were carried out to map nef CTL and Th epitopes in mice. A single CTL epitope and a dominant Th epitope, both restricted by H-2b, were identified. Consequently, C57BL/6 mice were immunized with different nef constructs by DNA 5 immunization means, and splenocytes from immunized mice were determined for Nef-specific CTL and Th responses using Elisopt assay and the defined T cell epitopes. The results showed that tpanef and tpanef(LLAA) were significantly more immunogenic than nef in terms of eliciting both CTL and Th responses.

Therefore, these aforementioned polynucleotides, when directly introduced 10 into a vertebrate *in vivo*, including mammals such as primates and humans, should express the respective HIV-1 Nef protein within the animal and in turn induce at least a cytotoxic T lymphocyte (CTL) response within the host to the expressed Nef antigen.

The present invention is not to be limited in scope by the specific 15 embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A pharmaceutically acceptable DNA vaccine, which comprises:
 - (a) a DNA expression vector; and,
 - 5 (b) a DNA molecule containing a codon optimized open reading frame encoding a Nef protein or immunogenic Nef derivative thereof, wherein upon administration of the DNA vaccine to a host the Nef protein or immunogenic Nef derivative is expressed and generates an immune response which provides a substantial level of protection against HIV-1 infection.
- 10 2. A DNA vaccine of claim 1 wherein the DNA molecule encodes wild type Nef.
- 15 3. A DNA vaccine of claim 2 wherein the DNA molecule contains the nucleotide sequence as set forth in SEQ ID NO:1.
4. The DNA vaccine of claim 3 which is V1Jns-opt nef (jrf1).
5. A DNA vaccine of claim 2 wherein the DNA molecule expresses a wild type Nef protein which comprises the amino acid sequence as set forth in SEQ 20 ID NO:2.
- 25 6. A DNA vaccine of claim 1 wherein the DNA molecule encodes an immunogenic Nef derivative which contains a nucleotide sequence encoding a leader peptide.
7. A DNA vaccine of claim 6 wherein the DNA molecule encodes an immunogenic Nef derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.
- 30 8. A DNA vaccine of claim 7 wherein the DNA molecule contains the nucleotide sequence as set forth in SEQ ID NO:3.
9. The DNA vaccine of claim 8 which is V1Jns-opt tpanef.

10. A DNA vaccine of claim 7 wherein the DNA molecule expresses an immunogenic Nef derivative which comprises the amino acid sequence as set forth in SEQ ID NO:4.

5

11. A DNA vaccine of claim 6 wherein the DNA molecule encodes an immunogenic Nef derivative modified at the dileucine motif of amino acid residue 174 and amino acid residue 175.

10 12. A DNA vaccine of claim 11 wherein the DNA molecule encodes an immunogenic Nef derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.

15 13. A DNA vaccine of claim 12 wherein the DNA molecule contains the nucleotide sequence as set forth in SEQ ID NO:7.

14. The DNA vaccine of claim 13 which is V1Jns-opt tpanef (LLAA).

15 15. A DNA vaccine of claim 11 wherein the DNA molecule expresses an 20 immunogenic Nef derivative which comprises the amino acid sequence as set forth in SEQ ID NO:8.

16. A DNA vaccine of claim 11 wherein the DNA molecule encodes a Nef protein where the glycine residue of amino acid residue 2 of Nef is modified to 25 encode for an amino acid residue other the glycine.

17. A DNA vaccine of claim 16 wherein the DNA molecule contains the nucleotide sequence as set forth in SEQ ID NO:5.

30 18. A DNA vaccine of claim 17 which is V1Jns-opt nef (G2A LLAA).

19. A DNA vaccine of claim 16 wherein the DNA molecule expresses an immunogenic Nef derivative which comprises the amino acid sequence as set forth in SEQ ID NO:6.

20. A DNA vaccine of claim 1 which further comprises an adjuvant.

21. A DNA vaccine of claim 20 whrerein the adjuvant is selected from the
5 group consisting of alumunum phosphate, calcium phosphate and a non-ionic block copolymer.

22. A pharmaceutically acceptable DNA vaccine, which comprises:

10 (a) a DNA expression vector; and,
(b) a DNA molecule containing an open reading frame encoding a Nef protein or immunogenic Nef derivative thereof, wherein upon administration of the DNA vaccine to a host the Nef protein or immunogenic Nef derivative is expressed and generates an immune response which provides a substantial level of protection against HIV-1 infection.

15

23. The DNA vaccine of claim 22wherein the DNA molecule expresses a wild type Nef protein which comprises the amino acid sequence as set forth in the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8.

20

24. A DNA vaccine of claim 22 which further comprises an adjuvant.

25. A DNA vaccine of claim 23 whrerein the adjuvant is selected from the group consisting of alumunum phosphate, calcium phosphate and a non-ionic block copolymer.

25

26. A method for inducing a cell mediated immune (CTL) response against infection or disease caused by virulent strains of HIV which comprises administering into the tissue of a vertebrate host a pharmaceutically acceptable DNA vaccine composition which comprises a DNA expression vector and a DNA molecule
30 containing a codon optimized open reading frame encoding a Nef protein or immunogenic Nef derivative thereof, wherein upon administration of the DNA vaccine to the vertebrate host the Nef protein or immunogenic Nef derivative is expressed and generates the cell-mediated immune (CTL) response.

27. The method of claim 26 wherein the vertebrate host is a human.

28. The method of claim 26 wherein the DNA vaccine is selected from the group consisting of V1Jns-opt nef (jrfl), V1Jns-opt tpanef, V1Jns-opt tpanef (LLAA),
5 and V1Jns-opt nef (G2A LLAA).

29. A substantially purified protein which comprises an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8.

10

1/10

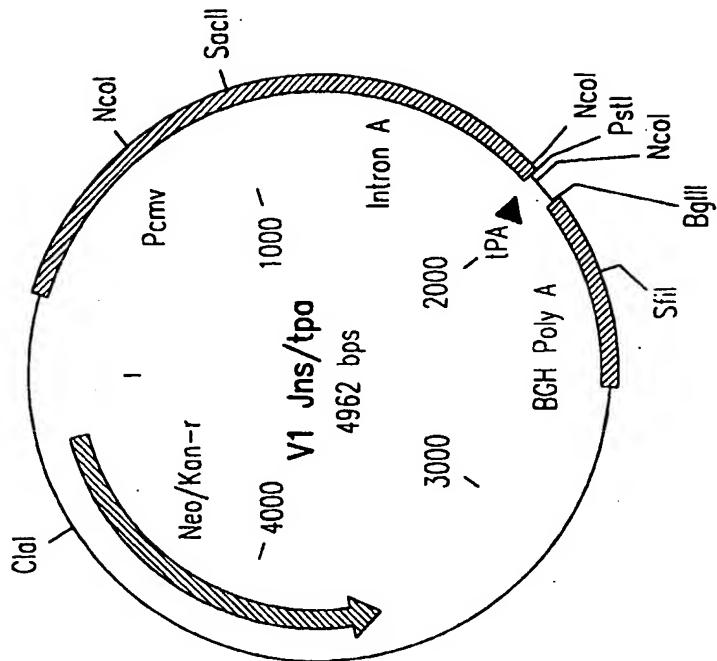


FIG. 1B

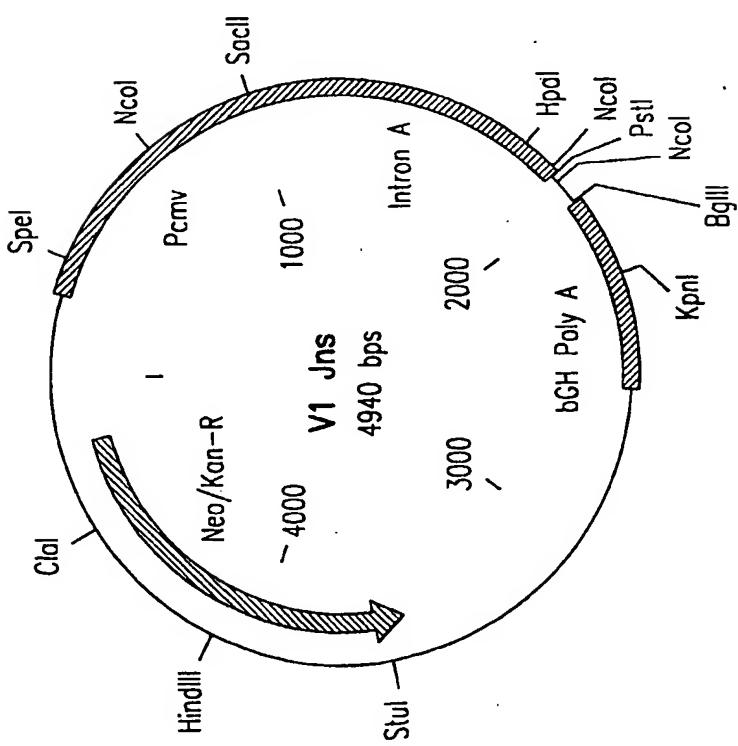


FIG. 1A

2/10

WT	- ATG GGT GGC AAG TGG TCA AAA CGT AGT GTG CCT GGA TGG TCT 	-42
OPT	- ATG GGC GGC AAG TGG TCC AAG AGG TCC GTG CCC GGC TGG TCC M G G K W S K R S V P G W S	-14
WT	- ACT GTA AGG GAA AGA ATG AGA CGA GCT GAG CCA GCA GCA GAT 	-84
OPT	- ACC GTG AGG GAG AGG ATG AGG AGG GCC GAG CCC GCC GCC GAC T V R E R M R R A E P A A D	-28
WT	- AGG GTG AGA CGA ACT GAG CCA GCA GCA GTA GGG GTG GGA GCA 	-126
OPT	- AGG GTG AGG AGG ACC GAG CCC GCC GCC GTG GGC GTG GGC GCC R V R R T E P A A V G V G V G A	-42
WT	- GTA TCT CGA GAC CTG GAA AAA CAT GGA GCA ATC ACA AGT AGC 	-168
OPT	- GTG TCC AGG GAC CTG GAG AAG CAC GGC GCC ATC ACC TCC TCC V S R D L E K H G A I T S S	-56
WT	- AAT ACA GCA GCT ACC AAT GCT GAT TGT GCC TGG CTA GAA GCA 	-210
OPT	- AAC ACC GCC GCC ACC AAC GCC GAC TGC GCC TGG CTG GAG GCC N T A A T N A D C A W L E A	-70
WT	- CAA GAG GAT GAG GAA GTG GGT TTT CCA GTC AGA CCT CAG GTA 	-252
OPT	- CAG GAG GAC GAG GAG GTG GGC TTC CCC GTG AGG CCC CAG GTG Q E D E E V G F P V R P Q V	-84
WT	- CCT TTA AGA CCA ATG ACT TAC AAG GGA GCT GTA GAT CTT AGC 	-294
OPT	- CCC CTG AGG CCC ATG ACC TAC AAG GGC GCC GTG GAC CTG TCC P L R P M T Y K G A V D L S	-98
WT	- CAC TTT TTA AAA GAA AAG GGG GGA CTG GAA GGG CTA ATT CAC 	-336
OPT	- CAC TTC CTG AAG GAG AAG GGC GGC CTG GAG GGC CTG ATC CAC H F L K E K G G L E G L I H	-112
WT	- TCA CAG AAA AGA CAA GAT ATC CTT GAT CTG TGG GTC TAC CAC 	-378
OPT	- TCC CAG AAG AGG CAG GAC ATC CTG GAC CTG TGG GTG TAC CAC S Q K R Q D I L D L W V Y H	-126
WT	- ACA CAA GGC TAC TTC CCT GAT TGG CAG AAC TAC ACA CCA GGG 	-420
OPT	- ACC CAG GGC TAC TTC CCC GAC TGG CAG AAC TAC ACC CCC GGC T Q G Y F P D W Q N Y T P G	-140

FIG.2A

3/10

WT	- CCA GGA ATC AGA TTT CCA TTG ACC TTT GGA TGG TGC TTC AAG 	-462
OPT	- CCC GGC ATC AGG TTC CCC CTG ACC TTC GGC TGG TGC TTC AAG P G I R F P L T F G W C F K	-154
WT	- CTA GTA CCA GTT GAG CCA GAA AAG GTA GAA GAG GCC AAT GAA 	-504
OPT	- CTG GTG CCC GTG GAG CCC GAG AAG GTG GAG GAG GCC AAC GAG L V P V E P E K V E E A N E	-168
WT	- GGA GAG AAC AAC TGC TTG TTA CAC CCT ATG AGC CAG CAT GGG 	-546
OPT	- GGC GAG AAC AAC TGC CTG CTG CAC CCC ATG TCC CAG CAC GGC G E N N C L L H P M S Q H G	-182
WT	- ATA GAG GAC CCG GAG AAG GAA GTG TTA GAG TGG AGG TTT GAC 	-588
OPT	- ATC GAG GAC CCC GAG AAG GAG GTG CTG GAG TGG AGG TTC GAC I E D P E K E V L E W R F D	-196
WT	- AGC AAG CTA GCA TTT CAT CAC GTG GCC CGA GAG CTG CAT CCG 	-630
OPT	- TCC AAG CTG GCC TTC CAC CAC GTG GCC AGG GAG CTG CAC CCC S K L A F H H V A R E L H P	-210
WT	- GAG TAC TAC AAG GAC TGC TGA (SEQ ID NO:30)	-651
OPT	- GAG TAC TAC AAG GAC TGC TAA (contained within SEQ ID NO:1) E Y Y K D C (SEQ ID NO:2)	-216

FIG.2B

4/10

V1Jns/nef
PstI
~~CATGGGTCTTTCTGCAGTCACCGTCCCTTGAGATCTGCCACC~~ ATG GGC AAG TGG TCC AAG AGG TCC GTG CCC . . .
 M G G K W S K R S V P . . .
 . . . CAC CCC GAG TAC TAC AAG GAC TGC TAA *SrfI BgIII*
 H P E Y K D C * (contained within SEQ ID NO:27)

FIG.3A

V1Jns/nef(G2A,LLAA)

PstI
~~CATGGGTCTTTCTGCAGTCACCGTCCCTTGAGATCTGCCACC~~ ATG GCC AAG TGG TCC AAG AGG TCC GTG CCC . . .
 M A G K W S K R S V P . . .
 . . . CAC CCC GAG TAC TAC AAG GAC TGC TAA *SrfI BgIII*
 H P E Y K D C * (contained within SEQ ID NO:28)

FIG.3B

V1Jns/tpanef & V1Jns/tpanef(LLAA)

PstI
~~CATGGGTCTTTCTGCAGTCACCGTCCCTTGAGATCTGCCACC~~ ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG
 M D A M K R G L C C V . . .
 CTG CTG CTG TGT GGA GCA GTC TTC GTT TCG CCC AGC GAG ATC TCC AAG AGG TCC GTG CCC . . .
 L L C G A V F S P S S K R S V P . . .

SrfI BgIII
 CAC CCC GAG TAC TAC AAG GAC TGC TAA *SrfI BgIII*
 H P E Y K D C * (contained within SEQ ID NO:8) (SEQ ID NO:29)

FIG.3C

5/10

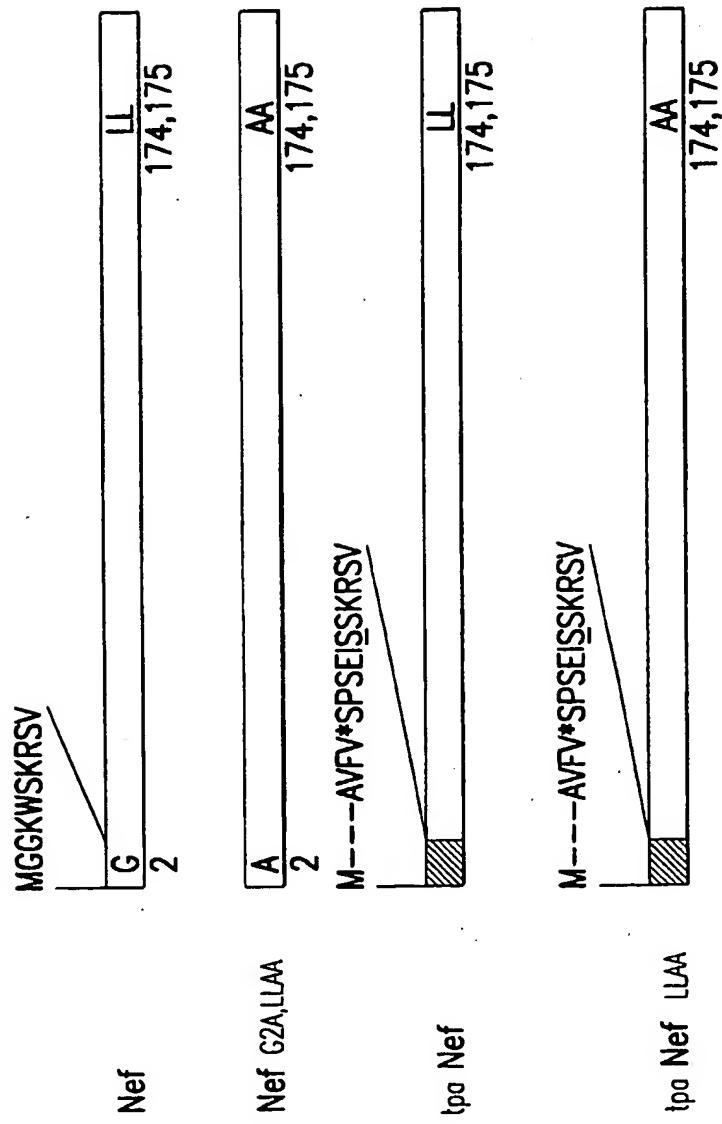


FIG. 4

6/10

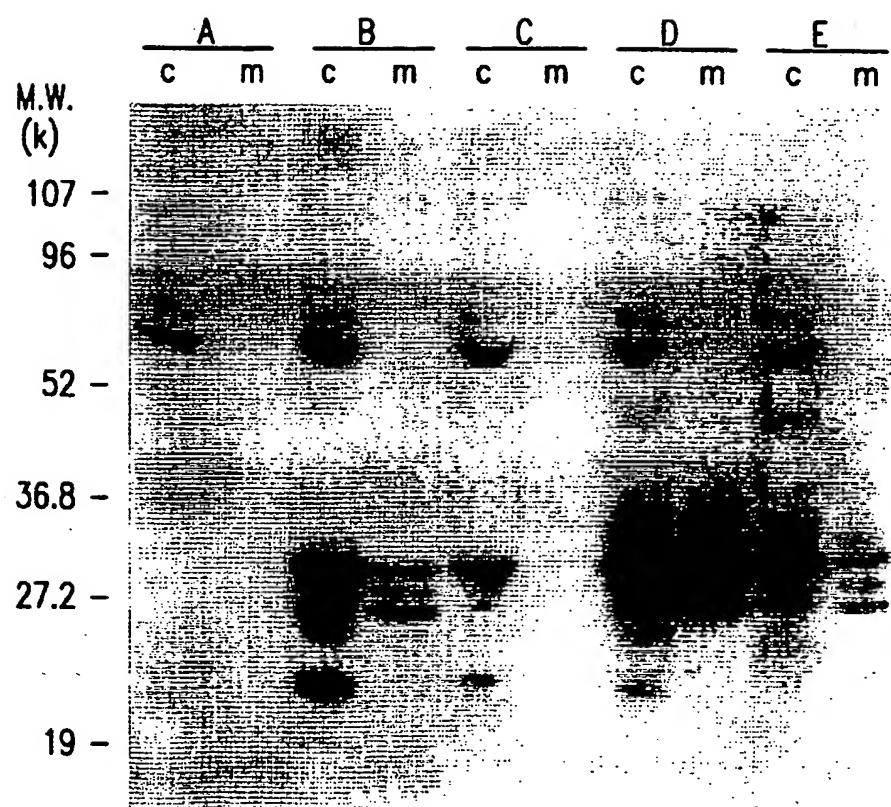


FIG.5

7/10

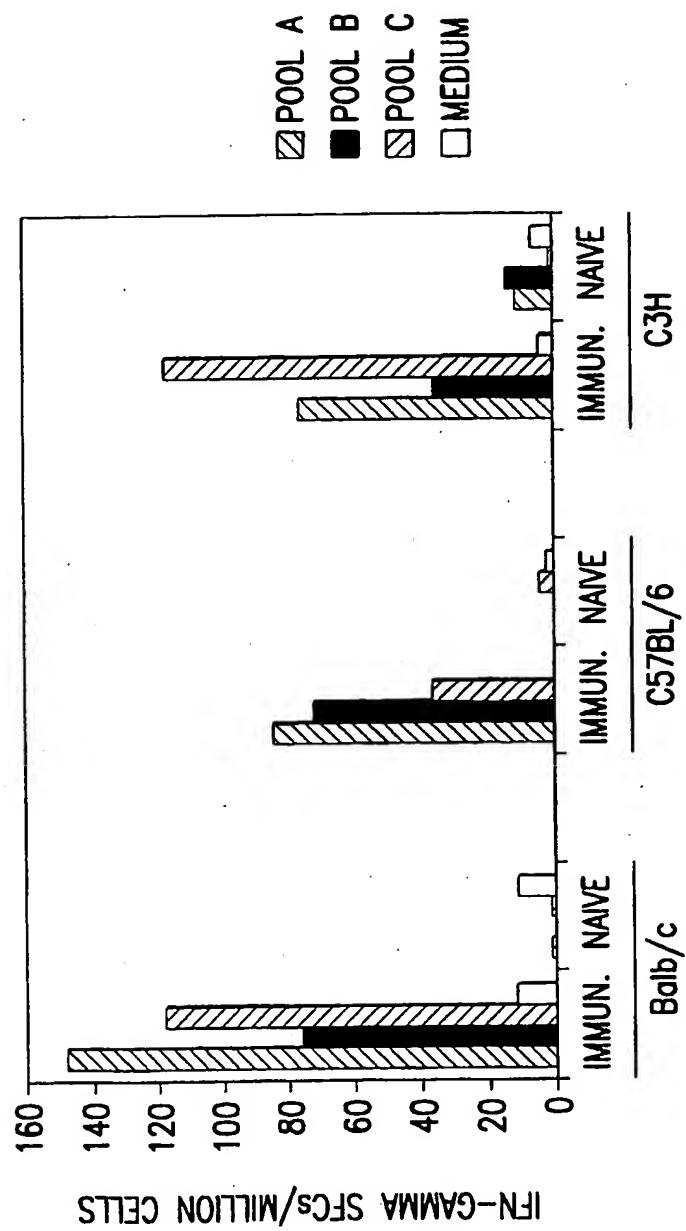


FIG. 6

8/10

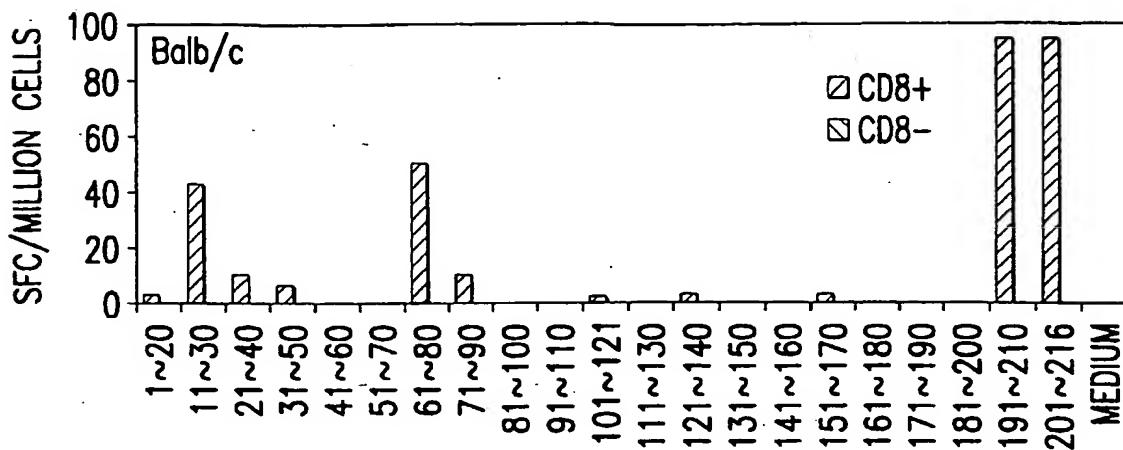


FIG.7A PEPTIDES

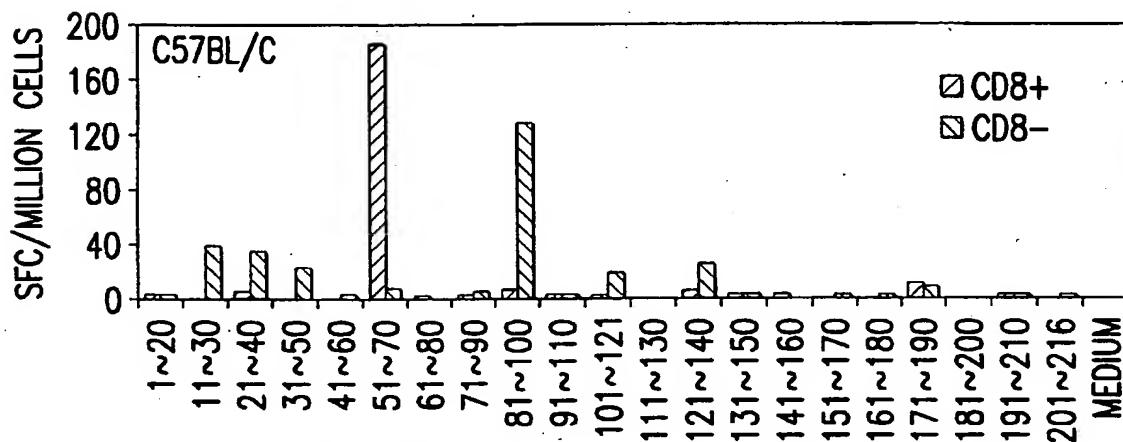


FIG.7B PEPTIDES

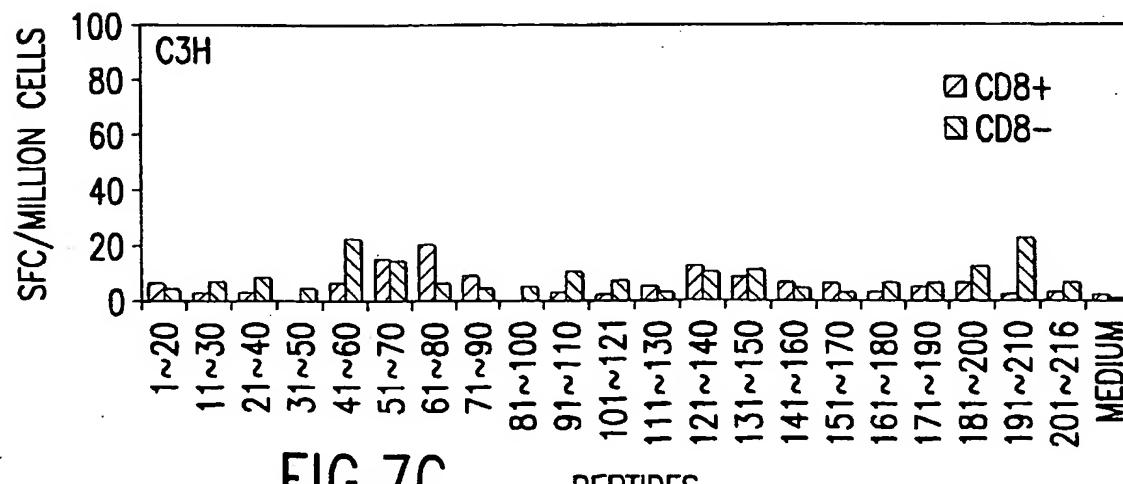


FIG.7C PEPTIDES

9/10

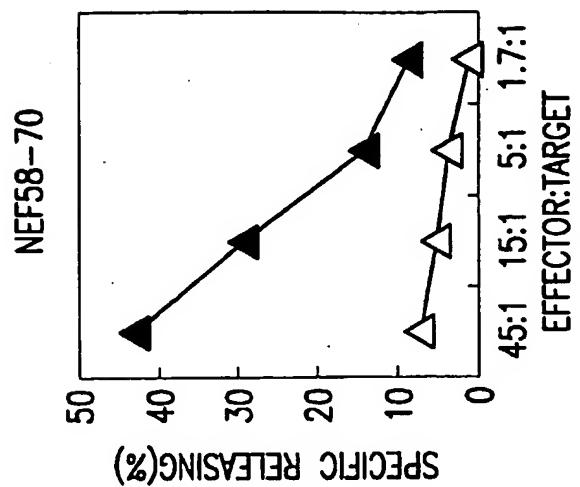


FIG.8C

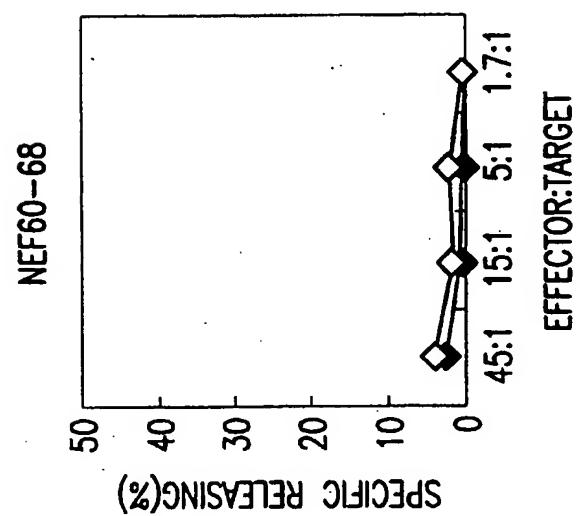


FIG.8B

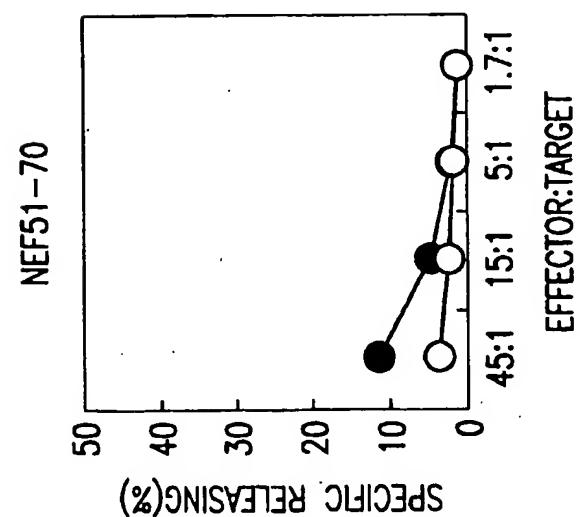


FIG.8A

10/10

EXPERIMENT 1

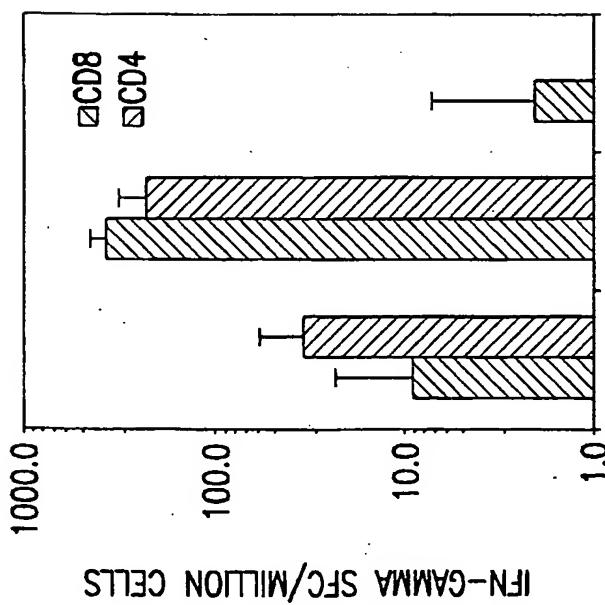


FIG.9A

ANTIGENS

NEF(G2A,LLM)
tppanef(LLM)
tppanef
NEF

EXPERIMENT 2

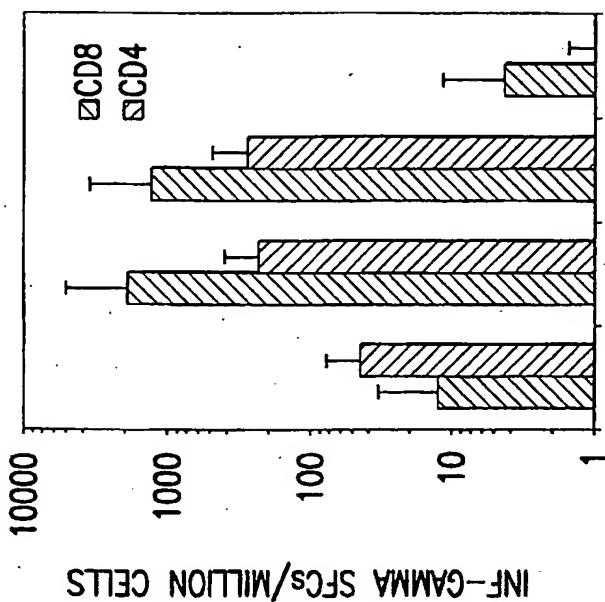


FIG.9B

ANTIGENS

NEF(G2A,LLM)
tppanef(LLM)
tppanef
NEF

SEQUENCE LISTING

<110> APPLICANT: Merck & Co., Inc.

<120> TITLE: POLYNUCLEOTIDE VACCINES EXPRESSING CODON
OPTIMIZED HIV-1 NEF AND MODIFIED HIV-1 NEF

<130> DOCKET/FILE REFERENCE: 20602Y

<160> NUMBER OF SEQUENCES: 30

<170> SOFTWARE: FastSEQ for Windows Version 4.0

<210> SEQ ID NO:1

<211> LENGTH: 671

<212> TYPE: DNA

<213> ORGANISM: Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (12) ... (662)

<400> SEQ ID NO:1

gatctgccac	c	atg	ggc	ggc	aag	tgg	tcc	aag	agg	tcc	gtg	ccc	ggc	tgg		50
Met	Gly	Gly	Lys	Trp	Ser	Lys	Arg	Ser	Val	Pro	Gly	Trp				
1	5								10							

tcc	acc	gtg	agg	gag	agg	atg	agg	gcc	gag	ccc	gcc	gcc	gac	agg		98
Ser	Thr	Val	Arg	Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Arg	
15	20								25							

gtg	agg	agg	acc	gag	ccc	gcc	gtg	ggc	gtg	ggc	gcc	gtg	tcc	agg		146
Val	Arg	Arg	Glu	Pro	Ala	Ala	Val	Gly	Val	Gly	Ala	Val	Ser	Arg		
30	35						40					45				

gac	ctg	gag	aag	cac	ggc	gcc	atc	acc	tcc	tcc	aat	acc	acc	ggc		194
Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr	Ala	Ala	Thr	
50	55									60						

aac	gcc	gac	tgc	gcc	tgg	ctg	gag	gcc	cag	gag	gac	gag	gag	gtg	ggc		242
Asn	Ala	Asp	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Asp	Glu	Glu	Val	Gly		
65	70								75								

ttc	ccc	gtg	agg	ccc	cag	gtg	ccc	ctg	agg	ccc	atg	acc	tac	aag	ggc		290
Phe	Pro	Val	Arg	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr	Tyr	Lys	Gly		
80	85								90								

gcc	gtg	gac	ctg	tcc	cac	ttc	ctg	aag	gag	aag	ggc	ggc	ctg	gag	ggc		338
Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	Gly	Leu	Glu	Gly		
95	100								105								

ctg	atc	cac	tcc	cag	aag	agg	cag	gac	atc	ctg	gac	ctg	tgg	gtg	tac		386
Leu	Ile	His	Ser	Gln	Lys	Arg	Gln	Asp	Ile	Leu	Asp	Leu	Trp	Val	Tyr		
110	115								120				125				

cac	acc	cag	ggc	tac	ttc	ccc	gac	tgg	cag	aac	tac	acc	ccc	ggc	ccc		434
His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr	Pro	Gly	Pro		

130	135	140	
ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg gtg ccc Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro 145	150	155	482
gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac aac tgc Val Glu Pro Glu Lys Val Glu Ala Asn Glu Gly Glu Asn Asn Cys 160	165	170	530
ctg ctg cac ccc atg tcc cag cac ggc atc gag gac ccc gag aag gag Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu 175	180	185	578
gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac gtg gcc Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala 190	195	200	626
agg gag ctg cac ccc gag tac tac aag gac tgc taa agcccgccc Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *	210	215	671

<210> SEQ ID NO:2

<211> LENGTH: 216

<212> TYPE: PRT

<213> ORGANISM: Human Immunodeficiency Virus - 1

<400> SEQ ID NO:2

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val 1	5	10	15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg 20	25	30	
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu 35	40	45	
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp 50	55	60	
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val 65	70	75	80
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp 85	90	95	
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His 100	105	110	
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln 115	120	125	
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg 130	135	140	
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro 145	150	155	160
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His 165	170	175	
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu 180	185	190	
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu 195	200	205	
His Pro Glu Tyr Tyr Lys Asp Cys 210	215		

<210> SEQ ID NO:3

<211> LENGTH: 719

<212> TYPE: DNA

<213> ORGANISM: Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (2)...(715)

<400> SEQ ID NO:3

c atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg ctg tgt gga	49
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly	
1 5 10 15	
gca gtc ttc gtt tcg ccc agc gag atc tcc tcc aag agg tcc gtg ccc	97
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro	
20 25 30	
ggc tgg tcc acc gtg agg gag agg atg agg agg gcc gag ccc gcc gcc	145
Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala	
35 40 45	
gac agg gtg agg agg acc gag ccc gcc gcc gtg ggc gtg ggc gcc gtg	193
Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val	
50 55 60	
tcc agg gac ctg gag aag cac ggc gcc atc acc tcc tcc aac acc gcc	241
Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala	
65 70 75 80	
gcc acc aac gcc gac tgc gcc tgg ctg gag gcc cag gag gac gag gag	289
Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu	
85 90 95	
gtg ggc ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac	337
Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr	
100 105 110	
aag ggc gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg	385
Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu	
115 120 125	
gag ggc ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg	433
Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp	
130 135 140	
gtg tac cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc	481
Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro	
145 150 155 160	
ggc ccc ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg	529
Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu	
165 170 175	
gtg ccc gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac	577
Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn	
180 185 190	
aac tgc ctg ctg cac ccc atg tcc cag cac ggc atc gag gac ccc gag	625
Asn Cys Leu Leu His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu	
195 200 205	
aag gag gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac	673

Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg			
15	20	25	
gtg agg agg acc gag ccc gcc gcc gtg ggc gtg ggc gcc gtg tcc agg			146
Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg			
30	35	40	45
gac ctg gag aag cac ggc gcc atc acc tcc tcc aac acc ggc gcc acc			194
Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr			
50	55	60	
aac gcc gac tgc gcc tgg ctg gag gcc cag gag gac gag gtg ggc			242
Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly			
65	70	75	
ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac aag ggc			290
Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly			
80	85	90	
gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg gag ggc			338
Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly			
95	100	105	
ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg gtg tac			386
Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr			
110	115	120	125
cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc ggc ccc			434
His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro			
130	135	140	
ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg gtg ccc			482
Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro			
145	150	155	
gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac aac tgc			530
Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys			
160	165	170	
gcc gcc cac ccc atg tcc cag cac ggc atc gag gac ccc gag aag gag			578
Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu			
175	180	185	
gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac gtg gcc			626
Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala			
190	195	200	205
agg gag ctg cac ccc gag tac tac aag gac tgc taa agccccggc			671
Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *			
210	215		

<210> SEQ ID NO:6

<211> LENGTH: 217

<212> TYPE: PRT

<213> ORGANISM: Human Immunodeficiency Virus - 1

<400> SEQ ID NO:6

Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val

1	5	10	15
Arg	Glu	Arg	Met
Arg	Arg	Met	Arg
Arg	Arg	Ala	Glu
Pro	Ala	Ala	Asp
Arg	Val	Arg	Arg
20	25	30	
Thr	Glu	Pro	Ala
Ala	Ala	Val	Gly
Gly	Val	Gly	Ala
Val	Ser	Arg	Asp
Leu	Glu		
35	40	45	
Lys	His	Gly	Ala
Ile	Thr	Ser	Ser
Asn	Thr	Ala	Ala
Thr	Asn	Ala	Asp
50	55	60	
Cys	Ala	Trp	Leu
Leu	Glu	Ala	Gln
Glu	Asp	Glu	Glu
Val	Gly	Phe	Pro
65	70	75	80
Arg	Pro	Gln	Val
Pro	Leu	Arg	Pro
Met	Thr	Tyr	Lys
Gly	Ala	Val	Asp
85	90	95	
Leu	Ser	His	Phe
Leu	Lys	Gly	Gly
Gly	Leu	Glu	Gly
Leu	Ile	His	
100	105	110	
Ser	Gln	Lys	Arg
Gln	Asp	Ile	Leu
Asp	Leu	Asp	Leu
Trp	Trp	Val	Tyr
115	120	125	
Gly	Tyr	Phe	Pro
Phe	Pro	Asp	Trp
Asp	Trp	Gln	Tyr
Tyr	Thr	Pro	Gly
130	135	140	
Phe	Pro	Leu	Thr
Leu	Thr	Phe	Gly
Trp	Cys	Phe	Cys
145	150	155	160
Glu	Lys	Val	Glu
Glu	Ala	Asn	Glu
Gly	Gly	Glu	Asn
165	170	175	
Pro	Met	Ser	Gln
His	Gly	Ile	Glu
Asp	Pro	Glu	Lys
Pro	Val	Lys	Glu
180	185	190	
Trp	Arg	Phe	Asp
Ser	Lys	Leu	Ala
Phe	His	His	His
195	200	205	
His	Pro	Glu	Tyr
Tyr	Lys	Asp	Cys
210	215		

<210> SEQ ID NO:7

<211> LENGTH: 720

<212> TYPE: DNA

<213> ORGANISM: Human Immunodeficiency Virus - 1

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (2)...(715)

<400> SEQ ID NO:7

c	atg	gat	gca	atg	aag	aga	ggg	ctc	tgc	tgt	gtg	ctg	ctg	tgt	gga
Met	Asp	Ala	Met	Lys	Arg	Gly	Leu	Cys	Cys	Val	Leu	Leu	Leu	Cys	Gly
1	5						10						15		

49

gca	gtc	ttc	gtt	tcg	ccc	agg	atc	tcc	tcc	aag	agg	tcc	gtg	ccc	
Ala	Val	Phe	Val	Ser	Pro	Ser	Glu	Ile	Ser	Ser	Lys	Arg	Ser	Val	Pro
20	25									30					

97

ggc	tgg	tcc	acc	gtg	agg	agg	atg	agg	agg	gcc	gag	ccc	gcc	gcc	
Gly	Trp	Ser	Thr	Val	Arg	Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala
35	40									45					

145

gac	agg	gtg	agg	agg	acc	gag	ccc	gcc	gtg	ggc	gtc	ggc	gcc	gtg	
Asp	Arg	Val	Arg	Arg	Thr	Glu	Pro	Ala	Ala	Val	Gly	Val	Gly	Ala	Val
50	55									60					

193

tcc	agg	gac	ctg	gag	aag	cac	ggc	gcc	atc	acc	tcc	aac	acc	gcc	
Ser	Arg	Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr	Ala
65	70									75			80		

241

gcc	acc	aac	gcc	gac	tgc	gcc	tgg	ctg	gag	gcc	cag	gag	gac	gag	gag
Ala	Thr	Asn	Ala	Asp	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Asp	Glu	Glu

289

85	90	95	
gtg ggc ttc ccc gtg agg ccc cag gtg ccc ctg agg ccc atg acc tac Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr 100	105	110	337
aag ggc gcc gtg gac ctg tcc cac ttc ctg aag gag aag ggc ggc ctg Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Leu 115	120	125	385
gag ggc ctg atc cac tcc cag aag agg cag gac atc ctg gac ctg tgg Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp 130	135	140	433
gtg tac cac acc cag ggc tac ttc ccc gac tgg cag aac tac acc ccc Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro 145	150	155	481
ggc ccc ggc atc agg ttc ccc ctg acc ttc ggc tgg tgc ttc aag ctg Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu 165	170	175	529
gtg ccc gtg gag ccc gag aag gtg gag gag gcc aac gag ggc gag aac Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn 180	185	190	577
aac tgc gcc gcc cac ccc atg tcc cag cac ggc atc gag gac ccc gag Asn Cys Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu 195	200	205	625
aag gag gtg ctg gag tgg agg ttc gac tcc aag ctg gcc ttc cac cac Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His 210	215	220	673
gtg gcc agg gag ctg cac ccc gag tac tac aag gac tgc taa Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys *	225	230	715
235			
agcccc			720
<210> SEQ ID NO:8			
<211> LENGTH: 237			
<212> TYPE: PRT			
<213> ORGANISM: Human Immunodeficiency Virus - 1			
<400> SEQ ID NO:8			
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly 1 5 10 15			
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser Lys Arg Ser Val Pro 20 25 30			
Gly Trp Ser Thr Val Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala 35 40 45			
Asp Arg Val Arg Arg Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val 50 55 60			
Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala 65 70 75 80			
Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu 85 90 95			
Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr 100	105	110	

Lys Gly Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu
 115 120 125
 Glu Gly Leu Ile His Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp
 130 135 140
 Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro
 145 150 155 160
 Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu
 165 170 175
 Val Pro Val Glu Pro Glu Lys Val Glu Ala Asn Glu Gly Glu Asn
 180 185 190
 Asn Cys Ala Ala His Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu
 195 200 205
 Lys Glu Val Leu Glu Trp Arg Phe Asp Ser Lys Leu Ala Phe His His
 210 215 220
 Val Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
 225 230 235

<210> SEQ ID NO:9

<211> LENGTH: 4945

<212> TYPE: DNA

<213> ORGANISM: E. coli

<400> SEQ ID NO:9

tcgcgcgttt	cgggtatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgttaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttata	ttggctcatg	300
tccaaacatta	ccgcacatgtt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcattt	gttcatagcc	catatatgg	gttccgcgtt	acataactt	cggttaatgg	420
ccgccttggc	tgaccggcca	acgacccccc	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggg	cttccattt	acgtcaatgg	gtggagtatt	tacggtaaac	540
tggccacttg	gcgactatc	aagtgtatca	tatgccaat	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgcc	ggcattatgc	ccagtagat	accttatgg	actttctac	660
ttggcagttac	atctacgtat	tagtcatcgc	tattaccatg	gtgatgcgtt	tttgcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccatatga	780
cgtcaatggg	agttttttt	ggcacaaaa	tcaacgggac	tttccaaat	gtcgtaacaa	840
ctccggccca	ttgacgcaaa	tgggggttag	gcgtgtacgg	tgggggtt	atataaggag	900
agctcggtta	gtgaaccgtc	agatcgcc	gagacgccc	ccacgctt	ttgacttcca	960
tagaagacac	cgggaccgtat	ccagcctcc	cggccggaa	cggtgcat	gaacgcggat	1020
tccccgtgcc	aagagtgtac	taagtaccc	ctatagactc	tataggcaca	cccccttggc	1080
tcttatgcat	gtataactgt	ttttggctt	gggcctatac	accccccgtt	ccttatgtca	1140
taggtgatgg	tatagcttag	cctatagtg	tgggttattt	accattattt	accactcccc	1200
tattggtgac	gatactttcc	attacta	cataacatgg	cttttgc	caactatctc	1260
tattggctat	atgccaatac	tctgtcc	agagactgac	acggactct	tattttaca	1320
ggatgggttc	ccatttatta	tttacaaatt	cacatataca	acaacccgt	cccccttggc	1380
cgcagttttt	attaaacata	gcgtgggatc	tccacgcgaa	tctcggtac	gtgttccgga	1440
catgggctct	tcccggtag	cggggagct	tccacatcc	acccctggc	ccatgcctcc	1500
agcggctcat	ggtcgtcgg	cagctcc	ctcctaacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtcc	cacaaggcc	tggcggtagg	gtatgttct	1620
gaaaatgagc	gtggagattt	ggctcgac	gctgacgc	atggaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagttt	gtattctgat	aagagtca	gttaactccc	1740
gttgcgtgc	tgttaacggt	ggagggc	gtatgtc	cagtact	tgctggcg	1800
cgcgcacca	gacataatag	ctgacagact	aacagactgt	tccttccat	gggtctttc	1860
tgcagtcc	gtccttagat	caccatggat	gcaatgaaga	gagggctct	ctgtgtgc	1920
ctgctgtgt	gagcgtt	cgtttcc	agcga	gatct	ctagttgc	1980
gccatctgtt	gtttggcc	ccccctgtcc	ttccttgc	ctggaagg	ccactccc	2040
tgtcc	taataaaaatg	aggaaatgc	atcgat	tgtgat	gtcattctat	2100
tctgggggtt	gggggtggggc	aggacagaa	ggggagg	tggaaagaca	atagcaggca	2160
tgcgtgggat	gggggtgggt	ctatggccgc	tgcggccagg	tgctgaagaa	ttgaccgg	2220
tcctcc	ccagaaaagaa	gcaggcacat	ccccctct	gtgacacacc	ctgtccacgc	2280

ccctggttct	tagttccagc	cccactcata	ggacactcat	agctcaggag	ggctccgcct	2340
tcaatccccac	ccgctaaagt	acttggagcg	gtctctccct	ccctcatcat	cccaccaaac	2400
caaaccttagc	ctccaaagagt	gggaagaaaat	taaagcaaga	taggctatta	agtccagagg	2460
gagagaaaaat	gcctccaaca	tgtgaggaag	taatgagaga	aatcatagaa	tttcttcgc	2520
ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	cggcagcgg	tatcagctca	2580
ctcaaaggcg	gtaatacgggt	tatccacaga	atcaggggt	aacgcagggaa	agaacatgtg	2640
agcaaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgttttcca	2700
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaga	ggtggcggaaa	2760
cccgacagga	ctataaaagat	accaggcggt	tcccccgtga	agctccctcg	tgcgtctcc	2820
tgttccgacc	ctgcccgtta	ccggataacct	gtccgccttt	ctcccttcgg	gaagcgtggc	2880
gctttctcat	agctcacgct	gttaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	2940
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	3000
tctttagtcc	aacccggtaa	gacacgactt	atgccactg	gcagcagcc	ctggtaacag	3060
gattagcaga	gcgaggtatg	taggcgggtc	tacagagttc	ttgaagtgg	ggcctaacta	3120
cggtcacact	agaagaacag	tatgggtat	ctgcgtctg	ctgaagccag	ttacccctcg	3180
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	3240
tgtttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	caagaagatc	ctttgatctt	3300
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tggcatgag	3360
attatcaaaa	aggatctca	cctagatct	tttaaattaa	aatgaagt	ttaaatcaat	3420
ctaaagtata	tatgataaa	cttggctcga	cagttacaa	tgcttaatca	gtgaggcacc	3480
tatctcagcg	atctgtctat	tgcgttcatac	catagttgc	tgactcgggg	ggggggggcg	3540
ctgaggctcg	cctcgtgaag	aagggtgtgc	tgactcatac	caggcctgaa	tcgc(ccc)atc	3600
atccagccag	aaagtgggg	agccacgggt	gatgagagct	ttgttgttag	tggaccagtt	3660
ggtgattttg	aacttttgc	ttgccacgga	acggtctcg	ttgtcgggaa	gatgcgtgat	3720
ctgatccctt	aactcagcaa	aagttcgatt	tattcaacaa	agccgcgc	ccgtcaagtc	3780
agcgtaatgc	tctgcccagt	ttacaaccaa	ttaaccaatt	ctgatttagaa	aaactcatcg	3840
agcatcaaat	gaaactgc当地	tttattcata	tcaggattat	caataccata	ttttgaaaaa	3900
agccgtttct	gtaatgaagg	agaaaactca	ccgaggcagt	tccataggat	ggcaagatcc	3960
tggtatcggt	ctgcgattcc	gactcgtcca	acatcaatac	aacctattaa	tttccctcg	4020
tcaaaaataa	ggttatcaag	tgagaaatca	ccatgagtga	cgactgaatc	cggtgagaat	4080
ggcaaaaagct	tatgcattt	tttccagact	tgttcaacag	gccagccatt	acgctcgtca	4140
tcaaaatcac	tcgcataac	caaaccgtt	ttcattcgt	attgcgcctg	agcgagacga	4200
aatacgcgt	cgctgttaaa	aggacaatta	caaacaggaa	tcgaatgca	ccggcgcagg	4260
aacactgcca	gcbcataac	aatatttca	cctgaatcag	gatattctc	taatacctgg	4320
aatgctgtt	tcccgggat	cgcagtgtg	agtaaccatg	catcatcagg	agtacggata	4380
aaatgcttga	tggtcggaaag	aggcataaat	tccgtcagcc	agtttagtct	gaccatctca	4440
tctgttaacat	cattggcaac	gctaccttgc	ccatgtttca	gaaacaactc	tggcgcatcg	4500
ggcttcccat	acaatcgata	gattgtcgca	cctgattgcc	cgacattatc	gcgagcccat	4560
ttatacccat	ataaatcage	atccatgttgc	gaatttaatc	cgccgcctcg	gcaagacgtt	4620
tcccggtgaa	tatggctcat	aacaccctt	gtattactgt	ttatgtaa	agacagttt	4680
attgttcatg	atgatatatt	tttatcttgc	gcaatgttac	atcagagatt	ttgagacaca	4740
acgtggcttt	cccccccccc	ccattattga	agcatttac	agggttatttgc	tctcatgagc	4800
ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttccgc	cacatttccc	4860
cgaaaagtgc	cacctgacgt	ctaagaaaacc	attattatca	tgacattaac	ctataaaaat	4920
aggcgtatca	cgaggccctt	tcgtc				4945

<210> SEQ ID NO:10
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:10
ctatataaagc agagctcggt tag

23

<210> SEQ ID NO:11
<211> LENGTH: 30

<212> TYPE: DNA
 <213> ORGANISM:Artificial Sequence

 <220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

 <400> SEQ ID NO:11
 gtagcaaaga tcttaaggacg gtgactgcag

30

<210> SEQ ID NO:12
 <211> LENGTH: 39
 <212> TYPE: DNA
 <213> ORGANISM:Artificial Sequence

 <220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

 <400> SEQ ID NO:12
 gtatgtgtct gaaaatgagc gtggagattg ggctcgcac

39

<210> SEQ ID NO:13
 <211> LENGTH: 39
 <212> TYPE: DNA
 <213> ORGANISM:Artificial Sequence

 <220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

 <400> SEQ ID NO:13
 gtgcgagccc aatctccacg ctcatttca gacacatac

39

<210> SEQ ID NO:14
 <211> LENGTH: 4432
 <212> TYPE: DNA
 <213> ORGANISM:E. coli

 <400> SEQ ID NO:14

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccc	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgtt	ctgaagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgtaaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacattttata	ttggctcatg	300
tccaaacatta	ccgccccatgtt	gacatttgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcattt	gttcatagcc	catatatgg	gttccgcgtt	acataactta	cggtaaatgg	420
cccgccctggc	tgaccgccc	acgacccccc	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaaataggga	cttccattt	acgtcaatgg	gtggagtatt	tacggtaaac	540
tgcccacttg	gcagttacatc	aagtgttatca	tatgccaagt	acgcccccta	ttgacgtcaa	600
tgacgtaaa	tggcccgcc	ggcattatgc	ccagttacatg	accttattgg	actttctac	660
ttggcagtac	atctacgtat	tagtcatcgc	tattaccatg	gtgatgcgtt	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccatgt	780
cgtcaatggg	agttttttt	ggcacaaaaa	tcaacgggac	tttccaaaat	gtcgttaacaa	840
ctccggccca	ttgacgcaaa	tggggcttag	gcgtgtacgg	tgggagggtt	atataagcag	900
agctcggtta	gtgaaccgtc	agatcgctg	gagacgccat	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggccggaa	cggtgcatgg	gaacgcggat	1020
tccccgtgcc	aagagtgcg	taagtaccgc	ctatacgttc	tataggccca	cccccttggc	1080
ttcttatgca	tgttatactg	tttttggctt	ggggctata	caccccccgt	tcctcatgtt	1140
atagggtatg	gtatagctta	gcctatacggt	gtgggttatt	gaccattatt	gaccactccc	1200
ctattggta	cgtatacttc	cattactaat	ccataacatg	gctctttgcc	acaactctct	1260
ttattggctt	tatgccaata	cactgtcctt	cagagactga	cacggactct	gtatttttac	1320
aggatgggtt	ctcattttt	atttacaaat	tcacatatac	aacaccacgg	tccccagtgc	1380
cccgagtttt	tattaaacat	aacgtggat	ctccacgcga	atctcggtt	cgtgtccgg	1440

acatgggctc ttctccggta gccccggagc ttctacatcc gagccctgc cccatgcctc
cagcgactca tggtcgctcg gcagctccc gtcctaaca gtggaggcca gacttaggc
cagcacgatg cccaccacca ccagtgtgcc gcacaaggcc gtggcgttag ggtatgtgc
tgaaaatgag ctgggggagc gggcttgcac cgctgacgca ttggaagac ttaaggcagc
ggcagaagaa gatcgaggca gctgagttgt tggttctga taagagtcag aggttaactcc
cgttgcgtg ctgttaacgg tggagggcag tggtagtctga gcaacttcg ttgctgccgc
gccccccacc agacataata gctgacagac taacagactg ttctttcca tgggttttt
ctgcagtcac cgtccttaga tctgctgtgc ttcttagttt ccagccatct gttgttgc
cctccccctg gccttccttg accctggaaag gtgcactcc cactgtcctt tcctaataaa
atgagggaaat tgcatcgcat tggctgatggta ggtgtcattt tattctgggg ggtgggtgg
ggcagcacag caagggggag gattggaaag acaatagcag gcatgctgg gatgcgtgg
gctctatggg taccagggtg ctgaagaatt gacccggttc ctcctgggccc agaaagaagc
aggcacatcc ccttcctgt gacacaccct gtccacgccc ctggttctta gttccagccc
cactcatagg acactcatag ctcaggaggg ctccgcctt aatcccaccc gctaaagtac
ttggagcgtt ctctccctcc ctcatcagcc caccaaacc aacctagctt ccaagagtgg
gaagaaatta aagcaagata ggctttaaag tgcagaggga gaaaaatgc ctccaacatg
tgaggaagta atgagagaaa tcatagaatt tcttcgctt ctcgctcac tgactcgctg
cgctcggtcg ttcggctcg gcgagcggta tcagctact caaaggcggta aatacggtt
tccacagaat caggggataa cgcagggaaag aacatgtgag caaaaggcaca gaaaaaggcc
aggaaccgta aaaaggccgc gttgctggcg ttttccata gctccgccc ccctgacgag
catcacaaaa atcgacgctc aagtccagg tggcgaaacc cgacaggact ataaagatac
caggcggttc ccccttggaaag ctccctcg tgcgtccttgc ttccgaccct gcccgttacc
ggataacctgt ccgcctttct cccttcggga agcgtggcgc ttctcaatg ctcacgctgt
aggatctca gttcgggtgt a ggtcgctcg tccaaagctgg gctgtgtgc cgaaccccc
gttcagcccg accgctgccc cttatccggta aactatcg ttcgatccaa cccggtaaga
cacgacttat cgccacttggc agcagccact ggtAACAGGA ttagcagagc gaggtatgt
ggcgggtgcta cagagttttt gaaagtgggg cctaaactacg gctacactag aaggacagata
tttggatctt ggcgtctgtc gaaaggcaggta accttcggaa aaagagttgg tagctttga
tccggcaaac aaaccaccgc tggtagcgg ggtttttt tttcgaaagca gcaaggatc
cgcagaaaaa aaggatctca agaagatctt ttgatctttt ctacgggggtc tgacgctc
tggAACGAAA actcacgtt aaggatttt gtcatgagat tatcaaaaag gatcttcacc
tagatccccc taaattaaaa atgaagttt aaatcaatct aaagtatata tgataaaact
tggctgaca gttaccaatg cttaaatcgat gaggaccta ttcagcgat ctgtcttatt
cggtcatcca tagtgtcctg actccccggc gtgtagatata ctacgatacg ggaggctta
ccatctggcc ccagtgtgc aatgataccg cgagacccac gtcacccggc tccagatata
tcagcaataa accagccagc cggaaaggccc gagcgcagaa gtggccttgc aactttatcc
gcctccatcc agtttattaa ttgttgcggg gaagcttagag taagtagttt gccagttaat
agtttgcgca acgttgc tttttttt cattgtacca ggcacgttgc tgcacgctc gtcgttgg
atggcttcat tcagctccgg tttccaaacca tcaaggcgag ttacatgatc cccatgttg
tgcaaaaaag cggttagctc cttcggtctt ccgatcggtt tcagaagtaa gttggccgca
gtgttacac tcatggttt ggcacgttgc cataattctc ttactgtcat gccatccgt
agatgcttt ctgtgactgg tgagttactca accaagtcat tctgagaata gtgtatgcgg
cgaccgagtt gctcttgc ggcgtcaata cgggataata cccgcacaca tagcagaact
ttaaaatgtc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttacc
ctgtttagat ccagttcgat gtaacccact cgtgcaccca actgatctt agcatctt
actttccacca gcgttctgg gtggcaaaa acaggaaaggc aaaaatgcgc aaaaaaggga
ataaggccga cacggaaatg ttgaataactc atactcttcc ttttcaata ttatgttgc
atttatcagg gttattgtct catgagcgga tacatattt aatgtattt gaaaaataaaa
caaatagggg ttccgcgcac atttcccgaa aaagtgcac ctgacgtcta agaaaccatt
attatcatga cattaaccta taaaatagg cgtatcaca gggcccttgc tc
4432

<210> SEQ ID NO:15
<211> LENGTH: 4864
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQ ID NO:15
tcgcgcgtt cggtgatgac ggtgaaaacc tctgacacat gcagctcccc gagacggtca 60
cagttgtct gtaaaggcgat gccgggagca gacaagcccg tcagggcgcg tcaggggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgtt ctgagagtgc 180
accatatgcg atatgtaaaata ccqcacagat qcgttaaggag aaaataccgc atcagattgg 240

cattggcca	ttgcatacgt	tgtatccata	tcataaatatg	tacatttata	ttggctcatg	300
tccaacatta	ccggcatgtt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcatta	gttcatagcc	cataatatgg	gttccgcgtt	acataactta	cggtaatgg	420
cccgcctggc	tgaccgcccc	acgaccccc	cccattgacg	tcaataatga	cgtatgtcc	480
catagtaacg	ccaataggga	cttccattg	acgtcaatgg	gtggagttt	tacggtaaac	540
tgcccactt	gcagtacatc	aagtgtatca	tatgccaatg	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgct	ggcattatgc	ccagttacatg	accttatggg	actttctac	660
ttggcagtac	atctacgtat	tagtcatcg	tattaccatg	gtgatgcgtt	tttggcagta	720
catcaatggg	cgtggatagc	ggtttactc	acggggattt	ccaagtctcc	accccattga	780
cgtcaatggg	agtttgttt	ggcacaaaa	tcaacgggac	tttccaaaat	gtcgtaacaa	840
ctccgcctta	ttgacgaaa	tggcggtag	gctgtacgg	ttggagggtct	atataagcag	900
agctcggtt	gtgaaccggtc	agatcgctg	gagacgccc	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagccctcg	cggccggaa	cggtgtcattt	gaacgcggat	1020
tccccgtgcc	aagagtgacg	taagtaccgc	ctataagatc	tataggccca	cccccttggc	1080
ttcttatgca	tgctatactg	ttttggctt	ggggctata	caccccccgt	tcctcatgtt	1140
atagggtatg	gtatagctt	gcctataggt	gtgggttatt	gaccattatt	gaccactccc	1200
ctattggta	cgatacttcc	cattactaat	ccataacatg	gcttttgc	acaactctct	1260
ttattggcta	tatgccaata	cactgtcctt	cagagactga	cacggactct	gtatttttac	1320
aggatggggt	ctcattttt	atttacaaat	tcacatatac	aacaccaccc	cccccaagtgc	1380
ccgcagttt	tattaaacat	aacgtggat	ctccacgcga	atctcggtt	cgttcccg	1440
acatgggctc	ttctccggta	gcggcgggac	ttctacatcc	gagccctgtt	cccatgcctc	1500
cagcgactca	tggtcgctcg	gcagctcctt	gtccctaaca	gtggaggcga	gacttaggca	1560
cagcacgatg	cccaccacca	ccagtggtcc	gcacaaggcc	gtggcggtag	gttatgtgc	1620
tgaaaatgag	ctcggggagc	gggcttgcac	cgctgacgca	tttggaaagac	ttaaggcagc	1680
ggcagaagaa	gatgcaggca	gctgagttgt	tgtgttctga	taagagtcag	aggttaactcc	1740
cgttgcgggt	ctgttaacgg	tggagggcag	tgtagtctga	gcagttactcg	ttgctgcgc	1800
gcgcgcacc	agacataata	gctgacagac	taacagactg	ttcctttca	tgggtctttt	1860
ctgcagtcac	cgtctttaga	tctgtctgtc	cttctagttt	ccagccatct	gttggtttgc	1920
cctcccccgt	gccttcttg	acccttggaa	gtgcaactcc	cactgtcctt	tcctaataaaa	1980
atgaggaat	tgcatcgcat	tgtctgagta	ggtgtcattt	tattctgggg	ggtgggggtgg	2040
ggcagcacag	caagggggag	gattggaaag	acaatagcag	gcatgctgg	gatgcgggtgg	2100
gctctatggg	taccagggt	ctgaagaatt	gacccgggtt	ctcctgggccc	agaaaagaagc	2160
aggcacatcc	cctttctgt	gacacaccc	gtccacgc	ctggtttctta	gttccagccc	2220
cactcatagg	acactcatag	ctcaggaggg	ctccgcctt	aatcccaccc	gctaaagtac	2280
ttggagcggt	ctctccctcc	ctcatcgacc	caccaaaacca	accctagctt	ccaagagtgg	2340
gaagaaatta	aagaagatata	ggcttattaa	tgcagaggga	gagaaaatgc	ctccaacatg	2400
tgaggaagta	atgagagaaa	tcatagaatt	tcttcgc	cctcgtc	tgactcgctg	2460
cgctcggtcg	ttcggtcg	gctcggtt	tcagtcact	aaacggcggt	aatacgggtt	2520
tccacagaat	caggggataa	cccttcgtt	tcagggaaag	aaacatgtgag	gcaaaaggcc	2580
aggaaccgta	aaaaggccgc	gttgctggcg	ttttccata	ggctccgc	ccctgcacgag	2640
catcacaaaa	atcgacgctc	aagttaggg	ttgcgaaaacc	cgacaggact	ataaaagatac	2700
caggcggttc	ccccctggaa	ctccctcg	cgctcttctg	ttccgaccct	ggcgcttacc	2760
ggataccgt	ccgccttct	cccttcgg	agcgtggc	tttctcaatg	ctcacgcgt	2820
aggtatctca	gttcgggt	ggtcgttgc	tccaaatcg	gctgtgtgc	cgaacccccc	2880
gttcagcccg	accgtctgc	cttattccgt	aactatcg	ttgagtccaa	cccggttaga	2940
cacgacttat	cgccactggc	agcagccact	ggtaacagga	ttagcagac	gaggtatgt	3000
ggcgggtgct	cagagttt	gaagtgggg	cctaactacg	gctacactag	aaggacagta	3060
tttggatct	gctctctgt	gaagccagg	accttggaa	aaagagttgg	tagcttgc	3120
tccggcaaa	aaaccaccgc	tggtagcggt	ggtttttt	tttgcagca	gcagattacg	3180
cgcagaaaaa	aaggatctca	agaagatct	ttgatcttt	ctacgggtc	tgacgctc	3240
tggaacgaaa	actcacgtt	agggatttt	gtcagatgt	tatcaaaaag	gatcttc	3300
tagatccctt	taaattaaaa	atgaagttt	aaatcaatct	aaagtatata	tgagtaaact	3360
tggtctgaca	gttaccaatg	cttaatcg	gaggcaccta	tctcagcgt	ctgtcttatt	3420
cgttcatcca	tagtgcctg	actccgggg	gggggggc	tgaggct	ctcgtgaaga	3480
aggtgttgc	gactcatacc	aggcctga	cgccccatca	tccagccaga	aagtgggg	3540
gccacgggt	atgagagctt	tgttgcgtt	ggaccagtt	gtgat	acttttgc	3600
tgccacgaa	cggctcggt	tgtcggtt	atgcgtgatc	tgatcc	actcagaaa	3660
agttcgattt	attcaacaaa	gccgcgttcc	cgtcaagtca	gcgtatgt	ctgccc	3720
tacaaccaat	taaccaattc	tgattagaaa	aactcatcg	gcatcaat	aaactgc	3780
ttattcatat	caggattatc	aataccat	ttttggaaaa	gccgttctg	taatgaagga	3840
aaaaactc	cgaggcagtt	ccataggatg	gcaagatct	ggtatcggtc	tgcgattccg	3900

actcgccaa	catcaataca	acctattaaat	ttcccctcg	aaaaataag	gttatcaagt	3960
gagaaatcac	catgagtgc	gactgaatcc	ggtgagaatg	gaaaagctt	atgcattct	4020
ttccagactt	gttcaacagg	ccagccatta	cgctcgcat	aaaatcact	cgcataacc	4080
aaaccgttat	tcattcgtga	ttgcgcctga	gcgagacgaa	atacgcgatc	gctttaaaa	4140
ggacaattac	aaacaggaat	cgaatgcaac	cggcgcagga	acactgccag	cgcataaca	4200
atatttcac	ctgaatcagg	atattctct	aatacctgga	atgctgttt	cccgggatc	4260
gcagtggta	gtaaccatgc	atcatcagga	gtacggataa	aatgcttgat	ggtcggaaaga	4320
ggcataaatt	ccgtcagcca	gtttagtctg	accatctcat	ctgtaacatc	attggcaacg	4380
ctaccttgc	catgtttcag	aaacaactct	ggcgcatacg	gcttccata	caatcgatag	4440
attgtcgac	ctgattgccc	gacattatcg	cgagccatt	tataccata	taaatcagca	4500
tccatgttgg	aatttaatcg	cggcctcgag	caagacgtt	cccgttgaat	atggctata	4560
acacccttg	tattactgtt	tatgtaaagca	gacagttta	ttgttcatga	tgatataattt	4620
ttatcttgt	caatgtaaaca	tcagagattt	tgagacacaa	cgtggcttc	cccccccccc	4680
cattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	4740
tagaaaaata	aacaaatagg	ggtccgcgc	acatttcccc	aaaaagtgcc	acctgacgto	4800
taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	gaggcccttt	4860
cgtc						4864

<210> SEQ ID NO:16
<211> LENGTH: 4867
<212> TYPE: DNA
<213> ORGANISM:E. coli

<400> SEQ ID NO:16						
tcgcgcgtt	cgtgtatgac	ggtggaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgt	ctgagagtgc	180
accatatgc	gtgtgaaata	ccgcacagat	gcgttaaggag	aaaatccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttata	ttggctcatg	300
tccaacatta	ccgccccatgtt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcatta	gttcatagcc	catatatgg	gttccgcgtt	acataactta	cgttggaaatgg	420
ccgcctggc	tgaccgcccc	acgacccccc	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggagtatt	tacgttaaac	540
tgcggacttgc	gcagttatc	aagtgtatca	tatgccaagt	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggccgcct	ggcattatgc	ccagttatgc	accttatggg	actttcctac	660
ttggcagttac	atctacgtat	tagtcatcgc	tattaccatg	gtatgtcggt	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagttcc	acccattga	780
cgtcaatggg	agttttttt	ggcacaaaaa	tcaacgggac	tttccaaaat	gtcgttaacaa	840
ctccgcucca	ttgacgcaaa	tggggcttag	gcgtgtacgg	tggggaggtct	atataaggcag	900
agtcgttta	gtgaaccgtc	agatgcctg	gagacgccc	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggccgggaa	cggtgcattt	gaacgcggat	1020
tccccgtgcc	aagagtgcg	taagtaccgc	ctatagactc	tataggcaca	cccctttggc	1080
tcttatgcat	gtctatactgt	ttttggcttg	gggcctatac	accccccgtt	ccttatgcta	1140
taggtgatgg	tatagcttag	cctataggtt	tggttattt	accattattt	accactcccc	1200
tattgggtac	gatactttcc	attactaattc	cataacatgg	cttttgc	caactatctc	1260
tattggctat	atgccaatac	tctgtccctc	agagactgac	acggactctg	tatttttaca	1320
ggatggggtc	ccattttata	tttacaaatt	cacatataca	acaacgcctgt	cccccggtcc	1380
cgcagtttt	attaaacata	gcgtgggatc	tccacgcgaa	tctcggtac	gtgttccgg	1440
catgggtct	tctccggtag	cggcggagct	tccacatccg	agccctggtc	ccatgcctcc	1500
agcggctcat	ggtcgctcgg	cagctccctg	ctccataacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtcccg	cacaaggccg	tggcgtttag	gtatgtgtct	1620
gaaaatgagc	gtggagattt	ggctcgacacg	gctgacgcac	atgaaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagttgtt	gtattctgtat	aagagtcaaga	ggtactccc	1740
gttgcgttgc	tgttaacgg	ggagggcagt	gtatgtcgat	cagtaactcg	tgctgcccgc	1800
cgcgcacca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggttttttc	1860
tgcagtccac	gtcccttagat	ctgtgtgc	ttctagttgc	cagccatctg	ttgtttgccc	1920
ctccccctgt	ccttccttga	cccttggaaagg	tgcactccc	actgtcttt	cctaataaaaa	1980
tgagggaaatt	gcatcgatt	gtctgagtag	gtgtcattct	attctggggg	gtgggggtggg	2040
gcaggacacg	aagggggagg	attggaaaga	caatagcagg	catgctgggg	atgcgggtggg	2100
ctctatggcc	gctgcggcca	ggtgctgaag	aattgaccgg	gttccctctg	ggccagaaaag	2160
aagcaggcac	atcccccttc	ctgtgacaca	ccctgtccac	gccccctgg	tttagtccca	2220

gccccactca	taggacactc	atagctcagg	agggctccgc	cttcaatccc	acccgctaaa	2280
gtacttggag	cggctctcc	ctccctcatc	agccccaccaa	accaaaccta	gcctccaaga	2340
gtgggaagaa	attaaagcaa	gataggctat	taagtgcaga	gggagagaaaa	atgcctccaa	2400
catgtgagga	agtaatgaga	gaaatcatag	aattttttcc	gcttcctcgc	tcactgactc	2460
gctgcgctcg	gtcggtcggc	tgccggcagc	ggtatcagct	cactcaaagg	cggtaataacg	2520
gttatccaca	aatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	2580
ggccaggaac	ctaaaaaagg	ccgcgttgct	ggcgttttc	cataggctcc	gccccctga	2640
cgagcatcac	gctcaagtca	gaggtggcga	aaccgcacag	gactataaag		2700
ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	ccctgcccgt	2760
tacccgatac	ctgtccgcct	tttcccttc	gggaaagcgtg	gcccgttctc	atagctcacg	2820
ctgttaggtat	ctcagttcgg	tgttaggtcg	tcgttccaag	ctgggctgtg	tgacgaacc	2880
ccccgttca	cccgaccgcgt	gcccgttate	cggtaactat	cgtttttttt	ccaaacccgggt	2940
aagacacgac	ttatcgccac	tggcaggcgc	cactggtaac	aggattagca	gaggcaggta	3000
tgttaggcgg	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaagaac	3060
agtatttttgt	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaaagag	ttggtagctc	3120
ttgatccggc	aaacaaaccca	ccgctgttag	cgggtttttt	tttgggttca	agcagcagat	3180
tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	ggctcgacgc	3240
tcagtggaac	aaaaactcac	gttaaggat	tttggtcatg	agattatcaa	aaagatctt	3300
cacctagatc	ttttaaattt	aaaaatgaag	ttttaatca	atctaaagta	tatatgagta	3360
aacttggtct	gacagttaacc	aatgcttaat	cagtgaggca	cctatctcag	cgatctgtct	3420
atttcggtca	tccatagttt	cctgactcgg	ggggggggggg	cgctgaggtc	tgccctcgta	3480
agaaggtgtt	gctgactcat	accaggcctg	aatcccccac	tcatccagcc	agaaagtgag	3540
ggagccacgg	ttgatgagag	ctttgttgc	gggggaccag	ttggtgattt	tgaacttttgc	3600
ctttgccacg	gaacggctcg	cggtgtcggg	aagatgcgtg	atctgtatct	tcaactcagc	3660
aaaagttcga	tttattcaac	aaagccccc	tcccgtaag	tcagcgtaat	gctctgccag	3720
tgttacaacc	attaaccaa	ttctgatttag	aaaaactcat	cgagcatcaa	atgaaactgc	3780
aatttattca	tatcaggatt	atcaatacca	tattttggaa	aaagccgtt	ctgtaatgaa	3840
ggagaaaaact	caccgaggca	gttccatagg	atggcaagat	cctggtatcg	gtctgcgatt	3900
ccgactcgctc	caacatcaat	acaacctatt	aattttccct	cgtcaaaaaat	aaggttatca	3960
agtgagaaat	caccatgagt	gacgactgaa	tccggtgaga	atggcaaaaag	cttatgcatt	4020
tctttccaga	cttggtaaac	aggccagcca	ttacgctcg	catcaaaatc	actcgcatca	4080
accaaaccgt	tatttattcg	tgattgcgcc	tgagcgagac	gaaatacgcg	atcgctgtta	4140
aaaggacaaat	tacaaacacgg	aatcgatgc	aaccggcgca	ggaacactgc	cagcgcatca	4200
acaatatttt	cacctgaatc	aggatattct	tctaataccct	ggaatgctgt	tttccggggg	4260
atcgcaatgg	tgatgtacca	tgcatacatca	ggagtacgga	taaaatgtt	gatggtcgga	4320
agaggcataa	atccgtcag	ccagtttagt	ctgaccatct	catctgttaac	atcattggca	4380
acgctaccctt	tgccatgttt	cagaaacaaac	tctggcgcat	cgggcttccc	atacaatcga	4440
tagattgtcg	cacctgattt	cccgacatta	tcggcgagccc	atttatacc	atataaatca	4500
gcatccatgt	tggatattaa	tcggcgcc	gagcaacgcg	tttcccggtt	aatatggctc	4560
ataacacccc	tttatttact	gtttatgtaa	gcagacagtt	tttattgttca	tgatgtatata	4620
tttttatctt	gtgcaatgt	acatcagaga	ttttgagaca	caacgtgggt	ttcccccccc	4680
ccccattatt	gaagcattta	tcagggttat	tgtctcatga	gcggatacat	atttgaatgt	4740
attttagaaaa	ataaacaaat	aggggttccg	cgcacatttc	ccccaaaagt	gccacactgac	4800
gtctaagaaa	ccattattat	catgacattta	acctataaaa	ataggcgtat	cacgaggccc	4860
tttcgtc						4867

<210> SEQ ID NO:17

<211> LENGTH: 78

<212> TYPE: DNA

<213> ORGANISM:Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:17

gatcaccatg gatgcaatga agagagggct ctgctgtgtg ctgctgctgt gtggaggcgt
cttcgttcc cccagcga

60

78

<210> SEQ ID NO:18

<211> LENGTH: 78

<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:18
gatctcgctg ggcgaaacga agactgctcc acacagcagc agcacacagc agagccctct 60
cttcattgca tccatggt 78

<210> SEQ ID NO:19
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM:Homo sapien

<400> SEQ ID NO:19
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1 5 10 15
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser
20 25

<210> SEQ ID NO:20
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:20
ggtacaaaata ttggctattg gccattgtcat acg 33

<210> SEQ ID NO:21
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:21
ccacatctcg aggAACCGGG tcaatttttc agcacc 36

<210> SEQ ID NO:22
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:22
ggtacagata tcggaaagcc acgttgtgtc tcaaaaatc 38

<210> SEQ ID NO:23
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:23
cacatggatc cgtaatgctc tgccagtgaa acaacc 36

<210> SEQ ID NO:24
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:24
ggtacatgt cacgtaaaa agatcaaagg atcttcttg 39

<210> SEQ ID NO:25
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:25
ccacatgtcg acccgtaaaa aggccgcgtt gctgg 35

<210> SEQ ID NO:26
<211> LENGTH: 4864
<212> TYPE: DNA
<213> ORGANISM:E. coli

<400> SEQ ID NO:26

tcgcgcgttt	cgggtatgac	ggtaaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaaaggat	gccgggagca	gacaagcccg	tcagggcgcg	tcaggggtg	120
ttggccgggt	tcggggctgg	cttaactatg	cgccatcaga	gcagattgtt	ctgagagtgc	180
accatatgcg	gtgtaaaaata	ccgcacagat	gcgtaaaggag	aaaataccgc	atcagattgg	240
ctatggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttt	ttggctcatg	300
tccaaacatta	ccgcccatttt	gacatttatt	atttactatg	tattaatgt	aatcaattac	360
ggggtcattt	gttcatagcc	catatatgg	gttccgcgtt	acataactta	cgttaaatgg	420
cccgccctggc	tgaccggcca	acgccccccg	cccatgtacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaaataggga	ctttccattt	acgtcaatgg	gtggagttt	tacgttaaac	540
tgcccactt	gcagttatc	aagtgtatca	tatgccaat	acgcccccta	ttgacgtcaa	600
tgacgtaaa	tggcccgctt	ggcatttatgc	ccagttatgc	acctttatgg	actttctac	660
ttggcactac	atctacgtat	tagtcatcgc	tattaccatg	gtgtgcgtt	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaaagtctcc	acccattga	780
cgtcaatggg	agtttttttt	ggcacaaaaa	tcaacgggac	tttccaaaaat	gtcgtacaa	840
ctccggccca	ttgacgcaaa	tgggggttag	gcgtgtacgg	tgggaggtct	atataagcag	900
agctcgttt	gtgaaccgtc	agatcgctg	gagacgccat	ccacgcgttt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggccggaa	cgggtcattt	gaacgcggat	1020
tcccccgtcc	aagagtgcg	taagtaccgc	ctatacgatc	tatagggccca	cccccttggc	1080
ttcttatgca	tgttatactg	tttttggctt	gggttctata	caccccccgt	tcctcatgtt	1140
ataggtgat	gtatagctt	gcctataggt	gtgggttatt	gaccatttt	gaccactccc	1200
ctatgggtga	cgtatactt	cattactaat	ccataacatg	gtcttttgc	acaactctt	1260
ttatggctt	tatgccaata	cactgtcctt	cagagactga	cacggactct	gtatttttac	1320
aggatgggt	ctcattttt	atttacaaat	tcacatatac	aacaccaccc	tccccagtgc	1380
ccgcagttt	tatcaaacat	aacgtggat	ctccacgcga	atctcgggtt	cgtgtccgg	1440
acatgggtc	ttctccggta	gcggcggagc	ttctacatcc	gagccctgtt	cccatgcctc	1500
cagcgtact	tgttcgtctg	gcagctccct	gtccttaaca	gtggaggcca	gacttaggca	1560
cagcgtact	cccaccacca	ccagtgcc	gcacaaggcc	gtggcggtt	ggtatgtgc	1620
tgaaaatgag	ctcggggagc	gggcttgac	cgctgacga	tttggaaagac	ttaaggcagc	1680
ggcagaagaa	gatgcaggca	gctgagttgt	tgtgttctga	taagagtca	aggttaactcc	1740

cgttgcggtg	ctgtaaacgg	tggagggcag	tgtagtctga	gcagtactcg	ttgctgccgc	1800	
gcgcgcacc	agacataata	gctgacagac	taacagactg	ttccttcca	tggttcttt	1860	
ctgcagtcac	cgtccttaga	tctgctgtgc	cttctagttg	ccagccatct	gttgttgcc	1920	
cctccccctgt	gccttccttg	accctgaaag	gtgccactcc	cactgtcctt	tcctaataaaa	1980	
atgagggaaat	tgcatcgcat	tgtctgagta	ggtgtcatc	tattctgggg	ggtgggggtgg	2040	
ggcagcacag	caagggggag	gattgggaag	acaatagcag	gcatgctggg	gatgcgggtgg	2100	
gctctatggg	tacccaggtg	ctgaagaatt	gaccgggttc	ctcctggggc	agaaagaagc	2160	
aggcacatcc	ccttcctgt	gacacacccct	gtccacgccc	ctggttctta	gttccagccc	2220	
caactcatagg	acactcatag	ctcaggaggg	ctccgccttc	aatcccaccc	gctaaagtac	2280	
ttggagcggt	ctctccctcc	ctcatcagcc	caccaaacc	aacctagcct	ccaagagtgg	2340	
gaagaatta	aagcaagata	ggctattaag	tgcagaggga	gagaaaatgc	ctccaacatg	2400	
tgaggaagta	atgagagaaa	tcatagaatt	tcttccgctt	cctcgctcac	tgactcgctg	2460	
cgctcgctcg	ttcggctg	gchgagcgta	tcagctact	caaaggcggt	aatacggtta	2520	
tccacagaat	caggggataa	cgcagaaag	aacatgtgag	caaaggccca	gcaaaaggcc	2580	
aggaaccgta	aaaaggccgc	gttgctggcg	ttttccata	ggctccgccc	ccctgacgag	2640	
catcacaaaa	atcgacgctc	aagtcaagg	tggcggaaacc	cgacaggact	ataaagatac	2700	
caggcggttc	ccccctggaa	ctccctcg	cgcttcctg	ttccgaccct	gccgcttacc	2760	
ggataccctgt	ccgccttct	cccttcgg	agcgtggcgc	tttctcaatg	ctcacgctgt	2820	
aggtatctca	gttcgggt	ggtcgttc	tccaagctgg	gctgtgtgca	cgaacccccc	2880	
gttcagcccc	acccgtcg	cttacccgt	aactatcg	ttgagtc	cccgtaaga	2940	
cacgacttat	cgccactggc	agcagccact	ggtaacagg	ttagcagagc	gaggatgt	3000	
ggcgggtc	cagagtctt	gaagtgg	cctaactac	gtcacactag	aaggacagta	3060	
tttggtatct	gcgtctgt	gaagccagg	acccctggaa	aaagagttgg	tagctttga	3120	
tccggcaaac	aaaccaccgc	ttgttagccgt	ggttttttt	tttgc	caagca	3180	
cgcagaaaaa	aaggatctca	agaagatctt	ttgatcttt	ctacgggg	tgacgctcag	3240	
tggaaacaaa	actcacgtt	agggatttt	gtcatgagat	tatcaaaaag	gatcttcacc	3300	
tagatccctt	taaattaaaa	atgaagttt	aatcaatct	aaagtatata	tgat	aaact	3360
tggtctgaca	gttaccatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	3420	
cgttcatcca	tagttcg	actccgggg	ggggggggc	tgagg	tctgc	ctcgtaaga	3480
aggtgttgct	gactcatacc	aggcctgaat	cgccccatca	tccagccaga	aagtgg	3540	
gccacggtt	atgagagctt	ttttgttagt	ggaccagtt	gtgat	tttga	actttgctt	3600
tgccacggaa	cggctcg	tgtcggaa	atgcgtgatc	tgat	cctca	actcagcaa	3660
agttcgattt	attcaacaaa	gcccgg	cgtcaagt	gcgtatgt	ctgc	cagtgt	3720
tacaaccaat	taaccaattc	tgattagaaa	aactcatga	gcat	aaaaat	aaactgcaat	3780
ttattcatat	caggattatc	aataccat	ttttgaaaaa	gccgtt	ctg	taatgaagga	3840
gaaaactc	caggcagtt	ccataggat	gcaagatct	ggtat	cgg	tcgattccg	3900
actcgccaa	catcaataca	acattat	ttccctcg	aaaaataa	gttat	caagt	3960
gagaaatcac	catgagt	gactgaat	ggtgagaat	gcaaaag	ctt	atgcatttct	4020
ttccagactt	gttcaacagg	ccagccatta	cgctcg	caaaat	caact	cgcatcaacc	4080
aaaccgttat	tcattcg	ttgcgc	tg	gagacgaa	atc	cgatc	4140
ggacaattac	aaacaggaat	cgaat	cg	ggcgc	gagga	acactg	4200
atatttcac	ctgaat	cagg	atattctt	at	ac	cttgc	4260
gcagtgg	gt	accat	gatc	gt	at	tcgtt	4320
ggcataaaatt	ccgtc	agcc	gtttag	ct	at	tcgtt	4380
ctac	ttgc	catgtt	ccat	ctg	at	tcgtt	4440
attgtcgac	ctg	atttgc	gacat	ttc	caat	tcgtat	4500
tccatgttgg	aa	tttgc	tttgc	tttgc	tttgc	tttgc	4560
acacc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	4620
ttatctgttgc	caat	tttgc	tttgc	tttgc	tttgc	tttgc	4680
cattatgaa	gtat	tttgc	tttgc	tttgc	tttgc	tttgc	4740
tagaaaaata	aacaaatagg	gttccgc	acat	tttgc	tttgc	tttgc	4800
taagaaacca	ttattatcat	gacat	tttgc	tttgc	tttgc	tttgc	4860
cg	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	4864

<210> SEQ ID NO:27

<211> LENGTH: 139

<212> TYPE: DNA

<213> ORGANISM:E. coli / HIV-1

<400> SEQ ID NO:27

catgggtctt ttctgcagtc accgtccttg agatctgcca ccatggggcg caagtggcc

60

aagagggtccg tgccccaccc cgagtactac aaggactgct aaagcccggg cagatctgct	120
gtgccttcta gttgccagc	139
<210> SEQ ID NO:28	
<211> LENGTH: 139	
<212> TYPE: DNA	
<213> ORGANISM:E. coli / HIV-1	
<400> SEQ ID NO:28	
catgggtctt ttctgcagtc accgtccttg agatctgcca ccatggccgg caagtggcc	60
aagagggtccg tgccccaccc cgagtactac aaggactgct aaagcccggg cagatctgct	120
gtgccttcta gttgccagc	139
<210> SEQ ID NO:29	
<211> LENGTH: 203	
<212> TYPE: DNA	
<213> ORGANISM:E. coli / HIV-1	
<400> SEQ ID NO:29	
catgggtctt ttctgcagtc accgtcctta tatcttagatc accatggatc caatgaagag	60
agggcctctgc tgtgtgctgc tgctgtgtgg agcagtcttc gtttcgcccc gcgagatctc	120
ctccaagagg tccgtgcccc accccgagta ctacaaggac tgctaaagcc cgggcagatc	180
tgctgtgcct tctagttgcc agc	203
<210> SEQ ID NO:30	
<211> LENGTH: 651	
<212> TYPE: DNA	
<213> ORGANISM:Human Immunodeficiency Virus - 1	
<400> SEQ ID NO:30	
atgggtggca agtggtaaaa acgttagtgt cctggatggt ctactgttaag gaaaagaatg	60
agacgagctg agccagcagc agataggggtg agacgaactg agccagcagc agtaggggtg	120
ggagcagtat ctcgagacct gaaaaaacat ggagcaatca caagtagcaa tacagcagct	180
accaatgctg attgtgcctg gctagaagca caagaggatg aggaagtggg tttccagtc	240
agacctcagg tacctttaag accaatgact tacaaggggag ctgtagatct tagccactt	300
ttaaaaagaaa aggggggact ggaagggtca attcaactcac agaaaagaca agatatcctt	360
gatctgtggg tctaccacac acaagggtac ttccctgatt ggcagaacta cacaccagg	420
ccaggaatca gatttccatt gacctttgga tggtgcttca agcttagtacc agttgagcca	480
gaaaaggtag aagaggccaa tgaaggagag aacaactgct ttttacaccc tatgagccag	540
catggatag agacccgga gaaggaagtg ttagagtgg a gtttgacag caagctagca	600
tttcatcacg tgccccgaga gctgcatccg gactataca aggactgctg a	651